MAMMOGRAPHY SCREENING:

benefits, harms, and informed choice

Karsten Juhl Jørgensen

Doctoral dissertation

The Nordic Cochrane Centre
Rigshospitalet, Department 7811
Blegdamsvej 9
DK-2100 Copenhagen
Denmark
Phone: +45 35 45 71 51
Fax: +45 35 45 70 07
kj@cochrane.dk

Denne afhandling er af Det Sundhedsvidenskabelige Fakultet ved Københavns Universitet antaget til offentligt at forsvares for den medicinske doktorgrad. København, den 20/9 2012.

Ulla Wewer

Dekan

Forsvaret finder sted fredag den 11. januar 2013 klokken 14.00, CSS (tidl. Kommunehospitalet) lokale 1.1.18.

ISBN 978-87-90233-63-1

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my mentor and primary collaborator Peter C. Gøtzsche. His lectures in research methodology were the reason I entered science and he has supported me with great enthusiasm and inspiration from my first steps. I would also like to thank my co-authors John Brodersen, Ole Johannes Hartling, Anders Klahn, Jan Mæhlen, Margrethe Nielsen, and Per-Henrik Zahl. Kay Dickersin was incredibly generous and hosted me at the Johns Hopkins School of Public Health during part of my work.

Perhaps the most important contribution that made the efforts involved in this thesis such a pleasure is the working environment created by the incredibly dedicated staff at the Nordic Cochrane Centre: Olga Ahtirschi, Frihild Askham, Segun Bello, Irfan Dawood, Jannie Hedegaard, Asbjørn Hróbjartsson, Anders Jørgensen, Monica Kjellstrøm, Lasse T. Krogsbøll, Henrik Larsen, Andreas Lundh, Emma Maund, Rasmus Moustgaard, Jacob Riis (thank you for assistance with the graphic design of this thesis), Paolo Rosati, Jeppe Schroll, Britta Tendal, and Lasse Krogsbøll. I would like to especially thank Marian Pandal for her great personal interest in her colleagues and continuous efforts to make the Centre such a nice place to be. You are missed.

I would also like to thank my family, especially my wife Suresini, for putting up with me being mentally elsewhere (though physically present) for so long.

The Agnes and Poul Friis Fund kindly provided means to enable the completion of this thesis.

SUMMARY

The rationale for breast cancer screening with mammography is deceptively simple: catch it early and reduce mortality from the disease and the need for mastectomies. But breast cancer is a complex problem, and complex problems rarely have simple solutions.

Breast screening brings forward the time of diagnosis only slightly compared to the lifetime of a tumour, and screen-detected tumours have a size where metastases are possible. A key question is if screening can prevent metastases, and if the screen-detected tumours are small enough to allow breast conserving surgery rather than mastectomy.

A mortality reduction can never justify a medical intervention in its own right, but must be weighed against the harms. Overdiagnosis is the most important harm of breast screening, but has gained wider recognition only in recent years. Screening leads to the detection and treatment of breast cancers that would otherwise never have been detected because they grow very slowly or not at all and would not have been detected in the woman's lifetime in the absence of screening. Screening therefore turns women into cancer patients unnecessarily, with life-long physical and psychological harms. The debate about the justification of breast screening is therefore not a simple question of whether screening reduces breast cancer mortality.

This dissertation quantifies the primary benefits and harms of screening mammography. Denmark has an unscreened "control group" because only two geographical regions offered screening over a long time-period, which is unique in an international context. This was used to study breast cancer mortality, overdiagnosis, and the use of mastectomies. Also, a systematic review of overdiagnosis in five other countries allowed us to show that about half of the screen-detected breast cancers are overdiagnosed. An effect on breast cancer mortality is doubtful in today's setting, and overdiagnosis causes an increase in the use of mastectomies. These findings are discussed in the context of tumour biology and stage at diagnosis.

The information provided to women in invitations and on the Internet exaggerates benefits, participation is directly recommended, and the harms are downplayed or left out, despite agreement that the objective is informed choice. This raises an ethical discussion concerning autonomy versus paternalism, and the difficulty in weighing benefits against harms.

Finally, financial, political, and professional conflicts of interest are discussed, as well as health economics.

This thesis is based on the following papers:

1. Jørgensen KJ, Gøtzsche PC. Presentation on web sites of possible benefits and harms from screening for breast cancer: cross sectional study. *BMJ* 2004;328:148-51.
2. Jørgensen KJ, Gøtzsche PC. Content of invitations for publicly funded screening mammography. *BMJ* 2006;332:538-541.
3. Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark. A comparative study. *BMJ* 2010;340:c1241.
4. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
5. Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009;9:36.

INTRODUCTION

The debate over mammography screening has been one of the most heated and emotional in medicine over the past 30 years. It is not without cause that it has been termed "the mammography wars".[1] The discussions have been fuelled by several factors, prime amongst which is a strong wish among professionals, the public, and politicians to reduce mortality from breast cancer, but also economical and professional ambition. Careers are built on the success of mammography screening and the screening industry turns over 5 billion dollars each year in the United States alone, counting only the screening procedure itself.[2] Slow recognition of harms, improved understanding of cancer biology, and diverging views on the role of modern medicine regarding autonomy versus paternalism has made the debate multi-faceted and not simply a question of whether screening reduces breast cancer mortality.

Despite eight randomised trials including more than 600,000 women, there are still diverging views about the quantification of the benefits of mammography screening and screening recommendations, as seen in full flare after the 2009 update of the U.S. Preventive Services Task Force review[3,4,5], and the 2011 recommendations from The Canadian Task Force on Preventive Health Care.[6,7] But there is also increasing consensus that breast screening has important downsides.[8,9] Whether screening detects otherwise inconsequential cancer lesions (overdiagnosis) has been questioned, with claims that this does not happen at all.[10] But it is now widely recognised as a major problem, and even strong screening proponents that have previously considered overdiagnosis a small concern limited to *in situ* cases now acknowledge that it occurs for invasive cancers.[11,12] Overdiagnosis has been known as a problem at least from the 1980's[13] and the report from 2002 on mammography screening by the International Agency for Research on Cancer/World Health Organisation is very clear:

"*An obvious source of harm associated with any screening programme is unnecessary treatment of cancers that were not destined to cause death or symptoms.*"[14]

Mammography screening is the best-studied cancer-screening programme. Apart from the eight randomised trials, there have been numerous observational studies. Unfortunately, much more research effort has been devoted to explore the benefits than the harms, often using problematic surrogate outcomes in the observational studies, e.g. disease stage at the time of diagnosis as percentages in screen- versus non-screen detected cancers, rather than absolute numbers.[1]

The emphasis on the benefit in the scientific litterature is clearly reflected in the information offered to those invited.[15,16,17,18] In the information included with invitations to screening, there is often specific percentages indicating the expected reduction in breast cancer mortality, but relative risks are difficult to interpret. The most important harm (overdiagnosis) is usually not mentioned, and when it is, it is simply stated that it is uncertain how many that will be affected. This can be criticised on several accounts. First, the public has a right to be informed about the risks of health interventions and withholding information about important harms is illegal in several countries, and a violation of autonomy.[19] Second, it is questionable when a public authority directly recommends an intervention but feel uncertain about the quantification of the most important harm. Third, both the benefits and the harms were quantified in the randomised trials and uncertainties therefore also affect the estimate of both.

To evaluate screening, and medical interventions in general, it is not sufficient to establish if they reduce the risk of dying from a specific disease.[1,20] All important consequences must be known prior to implementation, also the negative ones. These must be weighted against each other, which cannot be done scientifically. It is a value judgement that does not have a "correct" answer. The question is if an avoided breast cancer death is more or less important than screening-induced, unnecessary cancer diagnoses. And what about screening-induced deaths from other causes? The best we can hope for is that a majority agrees whether screening should be offered. Once implemented, every individual has the right to make his or her own decision, without pressure to reach a certain conclusion and everyone should receive balanced, comprehensive information.

Over the past few years, several studies in major medical journals from various independent research groups have questioned the fundamental premises of breast screening.[21,22] Further, the lack of effect on breast cancer mortality we found in Denmark[23] has now been supported by others.[24,25] Also, our quantifications of overdiagnosis[26,27] have been supported by others, using different methods.[28,29,30] We have also shown that breast screening does not lead to less mastectomies because of overdiagnosis.[31,32] Moreover, our continuous

criticism of invitations to breast screening and official reports from screening programmes[17,18,33], and our exploration of conflicts of interest,[34] have contributed to the growing international concern about the intervention.

The debate reached a culmination with the announcement by Professor Sir Michael Richards that an independent assessment of the new evidence is to be performed by a panel of researchers in the United Kingdom who have not previously published in the field, and that the newly revised invitation to the National Health Service Breast Screening Programme (NHS BSP) will be re-written after just one year in service.[35,36] The research presented in this thesis has contributed importantly to these decisions.

BREAST CANCER MORTALITY

Tumour stages and screening theory

A reduction in breast cancer mortality is the primary goal of breast cancer screening. The fundamental idea is that the prognosis of an individual cancer may be changed from deadly to curable by detecting it earlier.[14] But as noted in a systematic review of breast cancer screening from the U.S. Preventive Services Task Force (U.S. PSTF), there is no direct evidence for this mechanism of effect.[37] Obtaining such evidence would require a study that compares a group of women treated immediately for their screen-detected breast cancer with a group treated some time after diagnosis. For obvious ethical and practical reasons, such a study has never been done. Lack of direct evidence for the mechanism of effect places high demands on the quality of the evidence for an effect.

The theory of improved prognosis through earlier detection is mainly based on clinical observation. Tumours that are small at the time of detection have a better prognosis than those detected when they are large and there is a linear correlation between tumour size at detection and the likelihood of metastasis.[38] It is tempting to conclude that if the large tumours were detected earlier, they would have the same favourable prognosis as those detected when small. But the importance of biological variation in the genetic constitution of tumours and the interaction with, for example, the host's immune defence system is becoming better understood.[39,40,41]

The tumours that are large at the time of detection may be the fast-growing, aggressive ones that are also biologically determined to be most likely to metastasise. Their prognosis may not be affected by earlier diagnosis as they may already have spread, regardless of screening. This problem is compounded by the fact that screening preferentially detects the slow-growing tumours with a long non-symptomatic phase (sojourn time), simply because there is more time to detect them. This is called length bias. Conversely, the fast-growing, aggressive cancers are more likely to present between screening rounds as interval cancers.[42] (Fig. 1).

A systematic review of the effect of breast screening on tumour size at detection found no reduction in the occurrence of tumours larger than 20 mm in diameter (a size often used to define advanced disease) in seven countries with breast screening operating for a long time.[21] Screening has caused large, persistent increases in the number of small

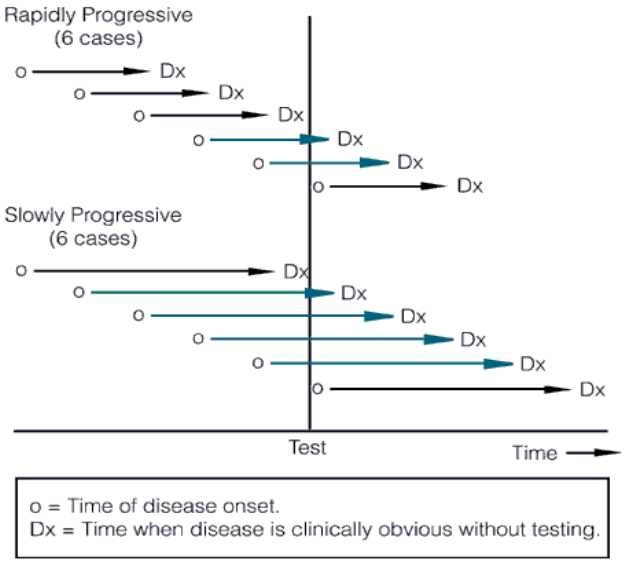

Figure 1. Length bias. From Welch 2004.[42]

invasive breast cancers and *in situ* lesions, as we[32] and others[8] have shown for the United States. But as this did not lead to a reduction in large cancers[21], we can conclude that these were not prevented by early detection. They "slipped through the screen" because they were fast-growing. Many of the small invasive cancers and *in situ* lesions that screening picked up were "extra" cancers. That is, they were overdiagnosed.[8,32]

It is not strange that fast-growing, aggressive cancers "slip through the screen", if we consider breast cancer growth and volume doublings of tumours (Fig. 2).[22] Screen-detected tumours are between 11 and 13 mm in diameter on average, whereas those detected in non-attenders and between rounds are about 22 mm on average.[43,44] This would correspond to 2 volume doublings out of the 32 necessary to reach 20 mm in diameter (Fig. 2). These numbers are from a modern-day setting but are not corrected for length bias, or small overdiagnosed cancers in the screened group (essentially extreme length bias), or self-selection bias due to non-attenders being different from attenders. The difference of about 10 mm is therefore an over-estimate of the true screening-induced reduction in tumour size. In the randomised trials, tumours in the control group were 21 mm on average[22], but this may have been reduced by opportunistic screening of about 25% of the women in some of the trials that published data on tumour size.[45] Tumours in the screened group were 16 mm on average[22], but this may also an underestimate due to overdiagnosed small cancers.

The difference in the trials corresponds to one volume doubling (Fig. 2).[22]

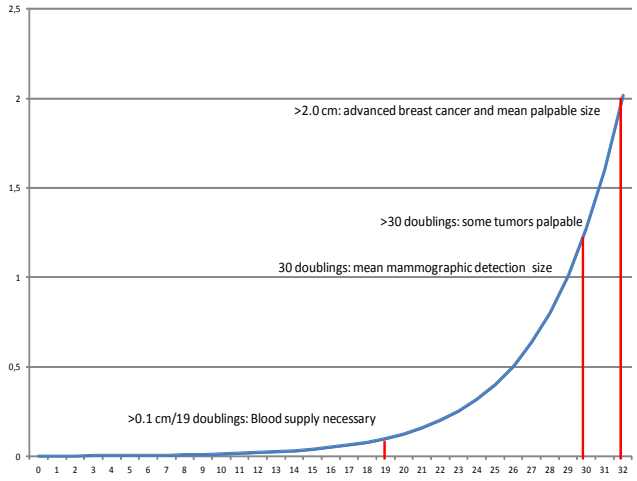

Figure 2: Tumour diameter (cm) versus volume doublings.

The mean tumour doubling time increases with age and was estimated at 233 days for women aged 50-59 years, and 260 days for women aged 60-69 years.[44] With screening intervals of 2-3 years, many tumours are missed and grow from a screen-detectable size of 10 mm to a size larger than 20 mm between screening rounds. As the fast-growing tumours double their volume much quicker, some in 50-100 days[44], most of them will not be detected with a screening interval of 1 year.

To allow continued growth, tumours require their own blood supply from the time they reach a size of about 1 mm in diameter, or 10^6 cells.[39,46] They can then spread through the bloodstream, if they possess the genetic constitution to form metastases. This is long before tumours are detected by screening, but still comparatively late in the total life cycle of a breast cancer (Fig. 2).

Studies using profiling of gene expression in breast cancer indicate that there are a small number of sub-classes, each with its own metastatic potential.[47] This potential is based on the expression of a large variety of genes and is inherent to the individual tumour. No studies have shown a change of sub-class with increasing size.[48]

It has also been shown that screen detection is a predictive factor independent from tumour size, as screen-detected tumours have a markedly better prognosis than clinically detected tumours of the same size.[48] These results were interpreted as an indication that screening preferentially detects cancers with a favourable prognosis, including overdiagnosed cancers. The study also indicated that the correlation between prognosis and tumour size at detection was only present for tumours over 1.3 cm in diameter at the time of detection. The reason that this relationship was not found for smaller tumours is likely the screening-induced "pollution" with small, overdiagnosed tumours. Increased sensitivity with technological development may therefore not be desirable.

Screening programmes for some other cancers are based on a fundamentally different principle and should be considered in their own right. Colorectal cancer screening aims to detect lesions that are not yet cancer, but polyps that may later become malignant. This reduces the problem of overdiagnosis of cancers and may even reduce colorectal cancer incidence.[49] Although there is still overdiagnosis of pre-cancer lesions, removing a polyp does not turn a healthy screenee into a cancer patient, nor does it require surgery with visible consequences, as does breast surgery. Such screening programmes can potentially constitute cancer prevention, whereas cancer screening based on early detection "creates" cancer patients through overdiagnosis and are arguably the opposite of prevention.

What we can learn from past experience

Several interventions for breast cancer have been introduced based on a theoretical mechanism of effect that seemed convincing at the time. Radical mastectomy was first line treatment until the late 1960's because cancer was considered to spread centrifugally through the tissue and lymphatic system from a primary lesion originating from a single cell.[40,50] It seemed logical that the more tissue that was removed around the primary lesion, the better the chance that all cancer cells would be eliminated. That patients still succumbed to breast cancer was attributed to lack of radicality and even more invasive surgery was thought to be the answer. Some women received excessively mutilating surgery.[40] Breast conserving surgery was considered inferior because it was not recognised that cancer can metastasise to distant organs through the bloodstream before it becomes clinically detectable and that metastases can re-surface after many years, even if the surgery removed the primary lesion and all affected lymph nodes. It was only when randomised trials showed that breast conserving surgery with adjuvant radiotherapy could provide similar survival rates as mastectomy that the

philosophy of "more radical surgery equals better survival" was abandoned.[40,50]

Recent randomised trials suggests that axillary dissection in early invasive breast cancer may do more harm than good, even in the presence of positive sentinel nodes.[51,52] Positive lymph nodes may be indicators of systemic spread, rather than the first line of defence against it, and in case of systemic spread, systemic treatment is needed.

High dose chemotherapy with bone marrow transplantation for advanced breast cancer gained wide support in North America in the 1990's and was also applied in some European centres.[53] The theory was that if some chemotherapy is good then a lot must be better. But chemotherapy not only kills cancer cells, it also knocks out the immune defence system and infections can pose a greater immediate threat than the cancer. To circumvent this limitation and hopefully kill all cancer cells, bone marrow was taken out prior to intensive chemotherapy, during which the patient was isolated in a near sterile environment. The bone marrow and immune defence system was reinstalled after chemotherapy.

There was great public and professional demand in North America to offer this treatment, despite lack of solid evidence. However, when randomised trials were finally done, partly because of pressure from health insurance agencies that had to pay for the expensive treatment, it was shown that the intervention was more harmful than beneficial. High-dose chemotherapy has toxic side effects and a sterile environment is difficult to maintain, leading to higher overall mortality.[54] To make matters worse, a positive trial from South Africa turned out to be fraud.[53]

Biology is often much more complex than it immediately appears. We must therefore require randomised trials that assess both the benefits and the harms before we implement health interventions.

Evidence from the randomised trials

Eight randomised trials of mammography screening have been performed, including more than 600,000 women.[45,55-63] It may seem surprising that there is still debate over the benefits and harms, but the results of the individual trials varies considerably and important biases contribute to the dispute.[45] Three comprehensive systematic reviews of all the trials have been undertaken.[6,37,45] In 2001, a Cochrane review concluded that methodological biases in the trials made the evidence for the intervention unreliable.[64] In 2002, a systematic review from the U.S. Preventive Services Task Force identified similar methodological problems in the trials as the Cochrane reviewers. They noted:

"The mortality benefit is small enough that biases in the trials could create or erase it."[37]

Despite the limitations, the Task Force evaluated that the trials were sufficiently reliable to conclude that mammography screening reduced breast cancer mortality by 16%, or that if 1224 women were screened, one death from breast cancer was prevented after 14 years of follow-up.[37] This is similar to the estimate in the most recent update of the Cochrane review, which included information about the trials published after the first review, and also a new trial, the Age-trial from the UK.[45,63] The updated Cochrane review considered it likely that mammography screening provides a relative risk reduction of 15%, or that 1 death from breast cancer is prevented for every 2000 women screened for 10 years.[45] A recent independent review by The Canadian Task Force for Preventive Health Care also reached similar estimates.[6]

A reduction of breast cancer mortality by 15-16% is about half the effect stated in invitations to mammography screening[16] and a reduction of 30-35% is also often claimed in the scientific literature.[34] The high estimates formed the basis for cost-effectiveness analyses and the decision to introduce national screening programmes, such as the NHS Breast Screening Programme (NHS BSP) in the UK, and the Danish breast screening programme.[13,65] According to the overview by the U.S. PSTF, such a large effect was only present in the Swedish Two-County trial and the Health Insurance Plan trial in New York.[37,58,60] These were the only two trials with published results in 1986 when the Forrest report paved the way for the NHS BSP.[13] The later trials showed effects between a 24% reduction[56] and a 2% increase in the relative risk of breast cancer mortality.[61,62]

The Two-County trial has been criticised for non-blinded outcome assessment.[45] When the cause of death was determined in the trial, the screening status of the women was known to the outcome assessor, which could influence the assignment and favour screening. A comparison of the published trial results with the official Swedish cause of death registry showed that several breast cancer deaths were lacking from the trial reports.[66] The publication of these results were vigorously opposed, resulting in an unfortunate example of poor editorial judgment with retraction of the original paper providing no

reason to the authors. The paper was later republished in another journal and the affair was described in the Lancet.[67,68] Whether the outcome assessment in the Two-County trial was blinded has been difficult to establish from publications and was investigated by the Pulitzer Prize winning journalist John Crewdson, who were able to get several testimonies from key investigators (though not from the lead investigator, László Tabár) that the outcome assessment was in fact not blinded.[69]

A recent publication re-assessed the causes of death[70] and found that the original outcome assessment fits official Swedish registry data. But the publication did not mention whether the new assessment was blinded and some of the authors were either primary investigators on the Two-County trial, or co-authors with these investigators on papers based on the original trial. This was not specified as conflicts of interest, and the choice of journal (Journal of Medical Screening) was also problematic, for reasons I will discuss later.

The New York Health Insurance Plan (HIP) trial was performed in the early 1960's when mammography equipment and breast cancer treatment were quite different from today.[58] Another shortcoming was that more women with a breast cancer diagnosed prior to the trial were excluded form the intervention arm than from the control arm, as women in the control arm were excluded based on unreliable registry data.[45] This would bias results in favour of the intervention.

In general, the later trials found smaller effects than the HIP and Two-County trials and the quality of the trials and their estimated effect on breast cancer mortality were inversely related; those that the Cochrane reviewers judged to be of good quality did not find much effect on breast cancer mortality, contrary to the trials of poor quality.[45] The U.S. PSTF judged the only the Canadian trials as being of "fair or better" quality, which the Cochrane reviewers classified as good (none were classified as good by U.S. PSTF).[37]

The U.S. PSTF did not quantify total mortality (deaths from any cause) in the trials. This outcome is not influenced by the assignment of cause of death and it also takes into account harms that lead to deaths. The Cochrane review found no impact on total mortality, regardless of the quality of the trials.[45] However, the trials did not include enough women to demonstrate an effect on this outcome, even if the intervention reduced the risk of dying from breast cancer by 30% and over 600,000 women participated. This is because the absolute benefit is very small; over a period of 10 years, about 10% of women aged 50-69 years would die from any cause, whereas only about 0.3% would die from a breast cancer detected within the same ten year interval (women diagnosed prior to the trial were generally excluded, as their outcome could not be affected). Although breast cancer is an important cause of death, the mortality from all other causes combined is much greater – 96-97 % of women will not die from breast cancer, but from something else. The chance of being "saved" by screening given a 33% reduction in risk is therefore 0.1% over 10 years, or 0.05% given a 15-16% reduction.

Expressing the effect as a relative risk reduction can be very misleading if it is not accompanied by information about the risk in absolute numbers. Essentially, a relative risk reduction of 33% does not indicate if the reduction is from 30% to 20%, 3% to 2%, or 0.3% to 0.2%. It is not surprising that invited women tend to overestimate the benefit[71-73], as they are only told about the relative risk reduction.[15,16]

Importantly, the trials could not demonstrate an effect on the total *cancer* mortality either (deaths from any cancer, including breast cancer), although they did include enough women.[45] The relative risk of death from any cancer in all the trials was 1.00 (95% CI 0.96-1.05), whereas a 29% reduction in breast cancer mortality should have resulted in a relative risk of 0.95. This is outside the 95% confidence interval (P=0.02).[45] There are two likely explanations: the reduction in breast cancer mortality has been overestimated due to bias; or mammography screening increases the mortality from other cancers (or both). Commonly used official statements such as "screening saves lives"[74] are therefore unsupported by the randomised trials. That mammography screening could increase mortality from other causes is related to overdiagnosis and subsequent overtreatment.

The age of the trials is a problem, particularly because of advances in treatment. Adjuvant therapy has improved survival substantially, also for women with metastases.[75] When fewer women die from their breast cancer because of better treatment, the number of women that screening can help is also reduced. As improved adjuvant therapy has benefited all prognostic groups[76], a synergistic effect of early detection and better treatment is unlikely.

Increased breast cancer awareness may have led to larger reductions in the average tumour size at

detection than screening. In 1978-9, the average tumour size at detection in Denmark was 33 mm, but this was reduced to 24 mm in 1988-89, a reduction of 9 mm before screening was introduced.[77] For comparison, the average difference in tumour size between the screened and non-screened groups in the trials was 5 mm, but this may be an overestimate due to overdiagnosis.[22] Such a difference corresponds to about 5% fewer tumours with metastases.[38] About 42% of tumours with an average size of 21 mm (such as those in the control arms of the trials) would have metastasised, on average. The reduction in tumour size caused by screening would therefore confer a relative risk reduction of (42%-5%)/42%=0.88, or 12% at most.[22] This mismatch between tumour biology and effect estimates in the trials indicate that the trials may have been biased in favour of screening. Data from current screening programmes is therefore vital to assess the effect today.

Evidence from observational studies

Observational studies based on individual patient history should not be used on their own to provide evidence for an effect of cancer screening because of the small effect and substantial biases.[1,14,78]

Such studies often receive considerable media-attention, but also criticism.[79-84] The fundamental problem is that many of these studies compare the outcome of screen detected and clinically detected cases in a setting where all are offered screening. This causes biases that favour the intervention, which has been known since the Forrest report:

"*It is not enough to compare the survival of patients with screen-detected cancers with the survival of those who present with symptoms. Although a longer survival of patients with screen-detected cancers might be observed, this alone is insufficient evidence that screening has prolonged survival because of various biases that may appear to enhance survival even if screening did not have an effect.*"[10]

Publications from public institutions that offer screening also make such comparisons. The Annual Review 2008 from the NHS Breast Screening Programme featured this headline:

"The 10-year fatality of screen-detected tumours is 50% lower than the fatality of symptomatic tumours."[74]

Stephen Duffy, Professor of Cancer Screening, is pictured next to the headline. No further explanation is provided. To a layperson, this is convincing evidence of an impressive effect of screening. But the fact is that the statement says nothing about the benefit of screening and is misleading because of three important biases. The first bias is the "Healthy Screenee Effect"[1], which refer to attendees being those with resources to worry about *potential* disease and do other things to improve their health.

"The screenees are the healthy, well-educated, affluent, physically fit, fruit and vegetable eating, non-smokers, with long-lived parents."[1]

They already have a comparatively good prognosis if they are diagnosed clinically, but are "selected" through their screening participation.

The considerable potential of selection bias to skew results was brilliantly illustrated by the authors of the Malmö trial.[85] They compared breast cancer mortality rates in participants versus non-participants within the screening arm of their randomised trial. After 9 years, by the end of 1986, the relative risk for breast cancer mortality was 0.96 (95% CI 0.68–1.35) when the trial was analysed as a randomised trial. But when the authors used a case–control design they found a significant (but false) 58% "effect" (OR matching for age; 0.42, 95% CI 0.22–0.78). Despite such clear evidence that the design is flawed, it is still used to evaluate screening[86], which I have criticized.[87]

The second bias is lead-time bias. The advancement of the time of diagnosis will improve the apparent survival time, even if screening does not make the women live longer in absolute terms (Fig. 3).[82,88]

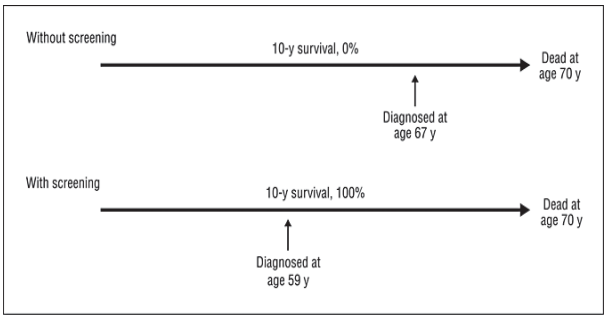

Figure 3. Lead-time bias. From Welch et al 2007.[82]

Third, length bias means that screening primarily detects the slow growing, least aggressive cancers and the screen-detected cases are therefore a select group with a fortunate prognosis (Fig. 1).[42] Fourth, overdiagnosis will introduce cancers that have an excellent prognosis because they would never have been fatal anyway, which artificially improves such statistics. All these biases were specified in the Forrest report in 1986.[13]

Danish observational studies

In 2005, a study reported a 25% reduction in breast cancer mortality in Copenhagen compared to unscreened regions in Denmark and a 37% reduction in breast cancer mortality among those who accepted the invitation to screening.[83] The study drew headlines such as "Cancer screening saves lives" in large Danish newspapers.[89] The reduction in breast cancer mortality was entirely attributed to screening mammography and the authors argued that differences in treatment between the regions were an unlikely confounding factor as there has been national treatment guidelines since the late 1970's. They disregarded that there are in fact substantial differences between regions, e.g. concerning the type of surgery used (mastectomy or breast conserving surgery). This has been highlighted by the Danish Breast Cancer Cooperative Group.[90] Such differences have led to monetary compensations for substandard care and pressure to centralise treatment.[91]

The study found that the *full* reduction in breast cancer mortality came already three years after screening in Copenhagen was implemented in 1991.[83] We criticised this[84] because it is incompatible with the randomised trials and screening theory.[14,92] When screening is introduced, the incidence increases reflecting both cancers that would have become symptomatic a few years later (earlier diagnoses) and overdiagnosed cases. However, the effect cannot occur until the time that the diagnosis was brought forward has passed. Further, if the diagnosis had been made clinically some time later, the patient would most likely have survived for some additional time. These two time periods must both pass before an effect of screening can occur. It also takes time from implementaion for all eligible women to be screened. In the randomised trials, an effect only began to emerge after 3-5 years with screening and the full effect was seen several years later still.[14,92]

Another problem with the 2005 study was the relatively few women that could benefit from screening in Copenhagen after three years. There were 45-86 breast cancer deaths per year during 1991 to 2006 in the age group that could potentially benefit (55-74 years), which consisted of about 50,000 women. In the first three years after screening was introduced, only some of these deaths would be from breast cancers also diagnosed within those first three years. Few could therefore have their prognosis affected by screening. And of those cancers that would both have been diagnosed and also killed the patient within those three years in the absence of screening, even fewer could have been catched by the screening programme, as it primarily detects the slow-growing lesions (length bias). This means that the conclusions in the study from 2005 were based on exceedingly few events.

It would have strengthened the conclusions if the authors had shown an identical effect in the other screened region in Denmark, Funen, which is about equally large. In Funen, however, the breast cancer mortality rates were similar to those in the non-screened areas throughout the observation period, both before and after screening (Fig. 4).

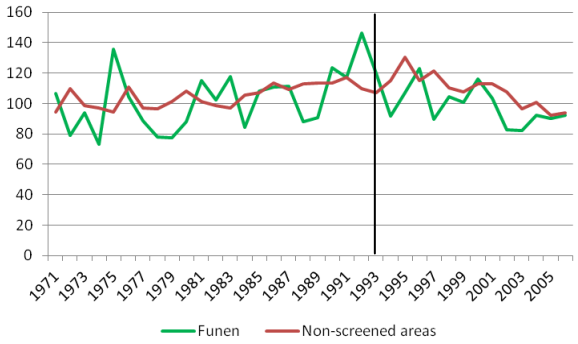

Figure 4. Breast cancer mortality per 100,000 women in Funen vs. non-screened areas in Denmark. Vertical line indicate when screening began in Funen.

This is despite markedly higher participation in Funen.[93] Some of the authors have later noted that Funen was not included in the 2005 study as they did not have 10 years of follow-up.[94] However, this would not have been necessary to document if the full effect had also occurred after three years in Funen.

While it might be true that the breast cancer mortality was 37% lower among those who actually attended screening relative to women in the non-screened areas[83], this does not mean that screening reduced mortality by 37%. The authors could not know which women that chose to attend. Again, the healthy screenee effect is at play.[1]

Modelling is sensitive to the choice of assumptions that the model is based on, e.g. the estimated average time that screening brings the diagnosis forward (lead time). Some of the same authors have later published calculations for Copenhagen using different models with different assumptions, with highly varying results.[95] Some results indicate an increase in breast cancer mortality in Copenhagen relative to the non-screened areas when screening

was introduced.[95] As no one knows which assumptions are correct, selecting which model to use is fraught with uncertainty.

What we found

We included data from both Copenhagen and Funen and found the same early reduction in breast cancer mortality in Copenhagen relative to the non-screened areas as in the 2005 study.[23,83] But the relative decline occurred well before screening could be of benefit and was only present in Copenhagen. In the period where screening could have the desired effect, breast cancer mortality in the non-screened areas was reduced at a rate of 2% per year versus 1% in the screened regions, although the decline started a few years later outside the screened regions.[23] We would have expected to see a more rapid decline in the screened areas, with an increasing difference in breast cancer mortality between screened and non-sceened areas over time. This did not happen. The greatest effect would be expected in the age group 55-74, shifted 5 years relative to the invited age group of women aged 50-69 years, as the effect would be delayed for the same reasons that the full effect could not occur in the first 5 years with screening (see above).

An even larger decline was seen in women who were too young to benefit from screening; 6% per year in the non-screened areas and 5% in the screened areas. The total decline in young women was also most pronounced in Copenhagen where it also started first. In Copenhagen, women too young to benefit from screening experienced a 60% decline in breast cancer mortality. We concluded that screening was unlikely to have caused a substantial reduction in breast cancer mortality in Denmark and that improved treatment offered a better explanation.[23]

It is possible that an effect was present, but too small to detect at population level. However, the expectation at the outset was that such an effect should be detectable. The Forrest Report noted that:

"This can be done approximately by examining trends in age-specific breast cancer mortality available from routine statistics."[13]

It was clear from both our study[23], and from a review of 30 European countries[96], that the effect of breast screening is too small to meet original expectations. The review found that the median change in breast cancer mortality was −37% (range −76% to −14%) in women under 50 years, −21% (−40% to 14%) in women aged 50-69 years, and −2% (−42% to 80%) in women over 70 year. To explore if a small effect is present, we will follow up our results and have requested individual patient data from the Danish National Board of Health.

Limitations

Assigning a cause of death is not simple, as anyone with experience in filling out deaths certificates will know, and screening could increase the number of deaths ascribed to the disease screened for. This has been called "sticky diagnosis bias"[97], as a diagnosis of a serious disease may follow a patient and influence decisions, also regarding the cause of death. Overdiagnosis would increase the number of women diagnosed with breast cancer which could "artificially" inflate mortality rates and lead to an underestimate of the screening effect. A counteracting bias is the "slippery linkage bias"[97]. Screening-induced deaths, for example from radiotherapy and chemotherapy in overdiagnosed healthy women, would not be ascribed to the screening intervention. The latter bias seems to have been more important in the randomised cancer screening trials, and this has strengthened the argument for using all-cause mortality as the primary effect measure.[97] The argument against this is that it requires large trials.

Hormone replacement therapy

With the publication of the results of the Women's Health Initiative trial in 2002,[98] and the Million Women Study in 2003,[99] the attitude towards hormone replacement therapy (HRT) changed abruptly. From a belief that HRT had a protective effect against breast cancer, it now appeared to increase both incident and fatal breast cancer. Shortly afterwards, the number of prescriptions fell in many Western countries[100]. A decline in the incidence of primarily hormone receptor positive breast cancer was observed in the United States beginning in mid-2002, reaching a plateau in 2004, which has been associated with the reduction in use of HRT since the 2002 Women's Health Initiative trial.[101] However, the conclusion was criticised in subsequent letters. The primary objections were that similar declines were absent in other countries that had reduced the use of HRT[100], and that the decline occurred too soon if the effect of HRT is *de novo* induction of breast cancers, rather than to stimulate growth of existing lesions[102]. Data from the United States (Fig. 5) shows that the increasing trend in incidence throughout the 1980's and 1990's changed already in 1998 while HRT use was peaking. Others have noted that the change in trend happened concurrently with declining participation in

mammography screening, from 78 % in 2000 to 72 % in 2005, particularly in women over 50 years which is the age group where the decline was also most pronounced.[103] The decrease in breast cancer incidence in 2002-4 is small compared to the increase associated with the introduction of breast screening (Fig. 5).

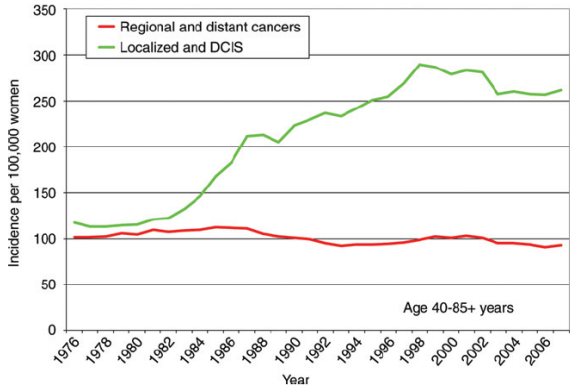

Figure 5. Incidence of breast cancer in the United States, regional/distant and localised/DCIS. From Jørgensen 2011.[32]

Other recent studies

Mette Kalager and colleagues[24] used the gradual introduction of breast screening in Norway to create historical screened and unscreened control groups, and a contemporary unscreened control group. They found that breast cancer mortality had declined in all age groups and regions since the 1990's. In the screened regions, the decline had been 10% larger than in the non-screened regions in the relevant age group, but the p-value was 0.13 (a confidence interval was not provided). A similar, also statistically non-significant, 8% difference was observed in the age group 70-84 years. The authors attributed this to the centralisation and specialisation of care that, due to governmental requirements, had happened simultaneously with the introduction of breast screening in the screened areas and benefited all age groups, leaving an effect of 2 percentage points to screening (the non-screened regions did not centralise care). However, there was a 4% (also statistically non-significant) difference in the *opposite* direction in the age group 20-49 years. The safest conclusion is that any difference in breast cancer mortality conferred by either screening or centralisation of treatment was too small to be detectable at population level.

The study has been criticised for its short follow-up (an average of 2.2 years after diagnosis) but this is a misunderstanding. The average follow-up was 6.6 years after the screening programme was introduced, which is how follow-up is defined in other studies and when an effect of breast screening emerged in the randomised trials[92]. Mette Kalager has now resigned from her position as Director of the Norwegian Breast Screening Programme, as she could not defend heading a screening programme that she would not participate in herself.[104]

Philippe Autier and colleagues compared breast cancer mortality rates in six neighbouring European countries: Sweden and Norway, Ireland and Northern Ireland, and the Netherlands and Belgium.[25] The idea was to compare demographically similar countries where one country had introduced breast screening in the early 1990's, while the other had introduced screening 10-15 years later. All compared countries had experienced equally large declines in breast cancer mortality, with the largest declines seen in young, unscreened women. The beginning of the declines in the screened age group was not related to the introduction of breast screening, often beginning long before it.

A third study from Turku, Finland, deserves mentioning, although it examined a slightly different question: the importance of the frequency of screening in women 40-49 years.[105] It was essentially a randomised design, with 14,765 women without breast cancer at age 40 years being assigned to either breast screening every year or every third year, based on their birth date (even or uneven date). This "unorthodox" randomisation method would not influence results in this case. As the authors note, practically all previous modelling studies, based on data from primarily the Two-County trial, indicate that young women would benefit particularly from more frequent screening, and that screening every 18 months is preferable. However, this study showed a relative risk for breast cancer mortality for triennial versus annual screening of 1.14 (CI: 0.59-1.27). That is, a trend in the *opposite* direction of that expected, albeit a small difference. More importantly, the relative risk for total mortality was 1.20 (CI 0.99-1.46), almost reaching significance. As the authors note, their study cannot determine if the difference between the two regimes is small, or if the programme as such "provided only a marginal effect overall at most" and that the study "points to the need for evaluating also the routine application of screening services".[105]

OVERDIAGNOSIS

Overdiagnosis is the detection of cancers through screening that would not have caused symptoms and therefore not have been detected in the lifetime of the woman in the absence of screening.[14] Because these cancers would never have posed a problem if there were no screening, their detection and treatment can only be harmful. It is sometimes referred to as inconsequential cancer diagnoses[1], although their detection has negative consequences.

Overdiagnosis represents the most important harm of screening and it has the potential to shift the balance between benefits and harms to the extent where screening is no longer justifiable. This has happened for other cancer screening programmes and is a likely cause of the opposition against the recognition of high levels of overdiagnosis in breast screening.

Competing causes of mortality and length bias

Although breast cancer is an important cause of death in middle aged and older women, it contributes with a comparatively small percentage of their total mortality, as the life time risk of dying from breast cancer is about 3-4% in most Western societies.[14] Screening programmes often operate with an interval of 1-3 years between rounds and primarily detects slow growing cancers while the fast growing cancers often become symptomatic and are detected between screening rounds.[42] Consequently, some women who had their slow growing breast cancer detected through screening will die from other causes before their cancers would have been diagnosed clinically. This can be considered a type of length bias (Fig. 1). This mechanism would be at play even if all breast cancers developed at the same rate and all had lethal potential, which is how breast cancer has been perceived historically. But there is large variation in the growth rate of breast cancers and some grow very slowly or not at all, and some even regress (Fig. 6).[106,107]

Some cancers are dormant and were not destined to cause symptoms in the lifetime of even long-lived individuals. Although these lesions fit all the usual pathological criteria of cancer, they behave quite differently and are sometimes called pseudo-disease.[42] But because of screening, these "cancers" are now detected and treated. The diagnosis and treatment of a pseudo-cancer cause the same physical and psychological harms as symptomatic cancers because it cannot be known if the individual cancer was overdiagnosed. Overdiagnosis is a major reason why screening for prostate cancer with prostate specific antigen (PSA) is so problematic.[108,109] It is also a major reason that we do not screen smokers for lung cancer with chest X-rays, the other reason being that it does not reduce lung cancer mortality.[110,111,112]

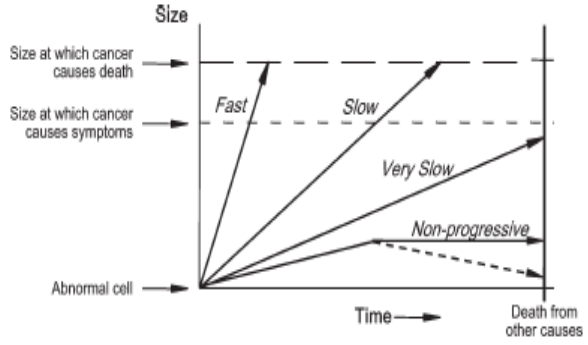

Figure 6. Variation in the growth of breast cancer. From Welch et al 2010.[112]

New lessons from prostate cancer screening

In 2009, the results from a European and an American randomised trial of prostate cancer screening with PSA have been published.[113,114] The European study was larger than the American study, including 162,243 and 76,693 men, respectively. The American study was handicapped because opportunistic PSA testing is common in the United States. This contaminated the control group and diluted any true effect, beneficial or harmful. The American trial did not show a reduction in the mortality from prostate cancer (relative risk 1.13, 95% CI 0.75-1.70). It did, however, show 22% overdiagnosis (relative risk 1.22, 95% CI 1.16-1.29).[114]

The European trial did not have much opportunistic screening of the control group and showed a 20% reduction in prostate cancer mortality (relative risk 0.80%, 95% CI 0.65-0.98).[113] Unsurprisingly, the level of overdiagnosis was much higher in the European trial. There was 71% overdiagnosis, with an incidence of 4.8% in the control group and 8.2% in the screened group. This translates into 48 unnecessary prostate cancer diagnoses for every life extended. Treatment for prostate cancer with surgery and radiotherapy is the most common approach and cause impotence in about 50% of cases, and also incontinence, although less commonly. The accompanying editorial noted that:

"*Serial PSA screening has at best modest effect on prostate-cancer mortality during the first decade of follow-up. This benefit comes at the cost of substantial overdiagnosis and*

overtreatment. It is important to remember that the key question is not whether PSA screening is effective but whether it does more good than harm. For this reason, comparisons of the [European trial] estimates of the effectiveness of PSA screening with, for example, the similarly modest effectiveness of breast cancer screening cannot be made without simultaneously appreciating the much higher risks of overtreatment associated with PSA screening."[108]

Overdiagnosis is more common in prostate cancer screening than in mammography screening because of the nature of the disease, the sensitivity of the blood test, and number of biopsies used (often 12 or more).[42] Slow growing, dormant invasive, and *in situ* prostate cancer lesions are common and autopsy studies have shown that such lesions are present in 60% of men in their 60's, whereas the lifetime risk of dying from prostate cancer is 3-4%.[42] The U.S Preventive Services Task Force have now issued a draft recommendation against routine screening with PSA.[115] A review of eight autopsy studies showed that there were also many undetected breast cancer lesions, both invasive and pre-invasive ones.[116]

Spontaneous regression of breast cancer

An ingenious study from Norway used the gradual introduction of screening in different administrative regions to show that women that were screened three times had 22% more cancers detected than women of the same age screened only once at the end of the observation period.[106] The original difference in incidence before the control population was screened was 57%. Extending the observation period so that one group was screened four times and the other group twice hardly impacted the difference, which was now 20%. This speaks against that the difference was due to limited sensitivity of mammography, as one would expect that almost all the "extra" breast cancers detected in the intensely screened population would also be detected in the control population when they were screened twice at the end. The authors concluded that the persistently higher incidence in the frequently screened group must have been due to cancers that would have spontaneously regressed in the absence of screening. An accompanying editorial acknowledged that this interpretation conflicts with how most lay people and clinicians perceive breast cancer, but also that other explanations for the observed difference in incidence had been dealt with and were less likely.[117] The study deservedly received considerable attention and is an important contribution to the way we perceive breast cancer.[118] The results have been supported by a similar, but stronger study from Sweden that included a much larger population, a wider age-range, and longer follow-up.[107] Also, the data were from a period where hormone replacement therapy use in the study population could have been only 4% at most, and 2% in the control population.[106]

A correlation between the number of screens and the number of cancers found supports a causal effect of screening to a greater extent than a simple correlation between time and event.[119]

Although spontaneous regression of invasive breast cancer may seem counter-intuitive, it has been described in the literature[120] and there is evidence from epidemiological studies that it occurs at population level.[121] The lack of more direct evidence may be due to the fact that practically all cases are treated and the natural course of sub-clinical breast cancers is largely unknown.[122] For neuroblastoma in children, screening caused a 100% increase in incidence.[123] We do not need long follow-up to determine that this was in fact extra, overdiagnosed cases as neuroblastomas are very rare in adults. The extra cases would therefore likely have regressed. This is supported by the clinical observation of spontaneous regression in all those 11 children that were diagnosed through screening, but where the parents chose a strategy of active monitoring.[42]

Quantifying overdiagnosis in mammography screening

It has been claimed that mammography screening can operate without overdiagnosis.[10] However, this is biologically impossible, as screening will inevitably detect cancers in women who die from other causes before their cancers would have become detected because of symptoms in the absence of screening.

Overdiagnosis in mammography screening is currently gaining wider acceptance as a significant problem[8,9,112], despite opposition from screening advocates.[124] Until recently, some screening proponents have claimed that if there were any overdiagnosis, it was confined to *in situ* lesions and that the problem was small.[11, 125-129] But in a recent study, which had many of the same authors, only overdiagnosis of invasive breast cancer was quantified, with an estimated ratio of two lives extended for every overdiagnosed case.[12] I had to ask the lead author on national British radio to learn that *in situ* cancers were not included in the study, as this was not mentioned in the study report.[12,130] I was also told that the reason *in situ* cancers were excluded was that overdiagnosis of such cases was

unimportant compared to invasive cancers. This is a major change of opinion. Overdiagnosis of invasive breast cancer is no longer possible to deny:

"*Twenty years ago the suggestion that pathology found in symptomless people might be inconsequential was greeted with derision. Now there are books published for the general public explaining the overdiagnosis problem.*"[1]

What remains to be established is its magnitude in a modern setting. The fundamental premise must be that any increase in the *lifetime* risk of breast cancer in a screened population compared to a non-screened population represents overdiagnosis.

The first quantification of overdiagnosis based on the randomised screening trials was published in 2000, but there were important biases in the trials that may have affected the estimate.[131] In some trials, there was opportunistic screening of the control group (e.g. one in four were screened in the control arm of the Malmö and Canadian trials) or the control group was screened at the end of the trial (e.g. the Two-County trial). There was also short duration of the randomised phase.[45] The trials are now getting old, and the technological development has increased sensitivity.

All of these biases would reduce estimates of overdiagnosis. But there are also biases in the opposite direction; the specialized staff in the trials may detect more cancer than in a public programme where it can be problematic to recruit skilled personnel, and a deliberately conservative attitude towards micro-calcifications and recalls in some programmes could also reduce overdiagnosis.[132]

Lead-time models

Estimating overdiagnosis in a clinical setting is important, but difficult. Often, lead-time models have been used[133], but they have important problems, as we have pointed out.[134,135] The basic premise of lead-time models is that screening causes a "shift" in the age-specific incidence due to the advancement of the time of diagnosis, which causes us to find those tumours we would otherwise have found some time into the future, in addition to those we would have found in the absence of screening. Thus, those aged e.g. 55 years will obtain the somewhat higher incidence of the age group a few years older in an unscreened population.

In the lead-time models, the expected increase in incidence is subtracted from the observed incidence increase in a screened population. Any remaining difference is considered overdiagnosis. The problem is that no one knows exactly how much breast screening advances the time of diagnosis and estimates have varied considerably, between 1 to 5 years.[129,136,137] As previously explained (Fig. 2), observed of tumour sizes at clinical and screen detection indicate that diagnosis is brought forward by one year at most, and likely considerably less.[22]

Too high estimates of lead time will over-compensate and lead to underestimates of overdiagnosis. Some have used the randomised trials to estimate lead-time[128] but do not consider overdiagnosis in the trials. Often, the stage at diagnosis is used to estimate how much screening had brought the diagnosis forward. But this can lead to substantial underestimates when screening causes overdiagnosis of early stage breast cancers.[135]

Using lifetime risk to quantify overdiagnosis

As we cannot differentiate between true and overdiagnosed cancers, overdiagnosis can only be defined statistically. Statistical models based on uncertain assumptions are problematic, but a different approach to estimate overdiagnosis in public screening programmes use the premise of an identical lifetime risk of breast cancer in a screened and a non-screened population in the absence of overdiagnosis.[26,27,138,139] Any excess incidence in the screened age group should be compensated by a reduction of the same number of breast cancers in women who pass the age limit for screening, as their cancers would already have been detected (Fig. 7). As there are less than one third as many women in the age group 70-80 years as in the age group 50-69 years, mainly because it is a narrower age range with increased total mortality, a compensatory decline measured per 100 000 women must be very large to compensate fully for the increase in younger women.

Figure 7: Model predicting breast cancer incidence with mammography screening. From Boer 1994.[139]

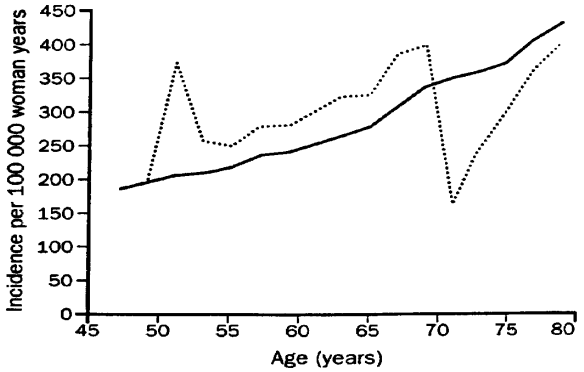

Figure: **Expected breast cancer incidence in 2-year age categories**
Solid line = not screened, dotted line = screened.

According to this model, excess incidence in the screened age group is expected, regardless if it is due to advancement of the time of diagnosis, overdiagnosis, or a combination of the two. If this excess incidence is not compensated by a decline in women who pass the age limit for screening, then screening has not advanced the time of diagnosis. Further, since the initial increase is required to indicate that advancement of the time of diagnosis has taken place, its absence, or the absence of a compensatory decline in older age groups, would mean that screening cannot have accomplished this goal and therefore cannot reduce mortality.

It also means that if an increase in incidence in the screened age group is not fully compensated by a subsequent decline in women who pass the age limit, any remaining difference is overdiagnosis.

The ideal way to quantify overdiagnosis would be a randomised trial with no contamination of the control group and life-long follow-up. Any difference in the total number of breast cancers between the screened and non-screened women would then be overdiagnosis. As such data does not exist, we must look at actual screening programmes.

What we did

In our systematic review, we quantified overdiagnosis using the premise of an unchanged lifetime risk of breast cancer in the absence of overdiagnosis.[26] To do this, it was essential to estimate what the background breast cancer incidence would have been in the absence of screening. The background incidence has been increasing steadily in most, but not all, Western populations in the years prior to screening.[140] Using a linear projection of the pre-screening trend, we quantified how much the incidence had increased in the screened age group compared to what was expected. We also quantified how much the incidence had fallen in older, previously screened women in relation to the background incidence, also projected from the pre-screening trend.[26]

To make reliable projections using linear regression, we required incidence data for at least seven years prior to the implementation of screening. We also required seven years of follow-up after the full implementation of screening to allow time for any compensatory drop in incidence among previously screened women to develop, and allow the incidence level in screened age groups to stabilise following the introduction of screening.[26] See Fig. 8 for an updated example.

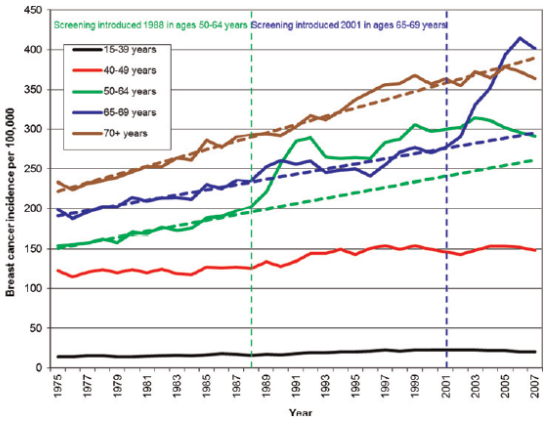

Figure 8. Breast cancer incidence in the UK. From Jørgensen 2011.[18] Data from Cancer Research UK.[141]

Some researchers have stated that we did not correct for the increasing background incidence[142], but this is not correct.[143] What may have confused some readers is that we excluded a small increase in incidence in the UK during the two years immediately prior to the roll-out of the NHS Breast Screening Programme, as we knew that a pilot screening programme had operated during this time.[13]

The remarkable consistency of the estimates of overdiagnosis between countries indicates that the data were trustworthy. Our search strategy did not include other databases than PubMed, but we also scanned reference lists and contacted authors. A control sample of all articles on breast screening published in 2004 used for another article[34] indicated that we had not missed any studies.

We were unable to find useable published data from Denmark, but Denmark offers a unique opportunity because two administrative regions that include 20% of the population have offered screening over seventeen years whereas the remaining regions have not.[27] The unscreened regions provide a "control group" and we were therefore able to evaluate if our projections of the expected development in background incidence were in accordance with actual observations from non-screened regions. We obtained detailed incidence data from the Danish National Board of Health and this allowed us to use Poisson regression analyses instead of simple linear regression to compensate for variation in age distribution. The results were very close to what we would have obtained with linear regression. These results indicate that the limitations we mentioned in our systematic review[26] (e.g. HRT use and demographic factors leading to a higher increase in

background incidence rates than projected) were likely unimportant.

What we found

Overdiagnosis in public mammography screening is an even greater problem than estimated from the randomised trials.[26,45] There was little or no compensatory decline in breast cancer incidence among previously screened women, despite long follow-up. We consistently found persisting, large increases in breast cancer incidence among screened women and that this increase was not present in other age groups. Our meta-analysis included data from five countries and we demonstrated that public mammography screening results in 52% overdiagnosis.[26] Currently, about one third of breast cancers are detected between screening rounds (interval cancers)[144] and our results therefore indicate that half the screen-detected cancers are overdiagnosed when *in situ* lesions are included (150% breast cancers in screened women compared to 100% in non-screened women, and one third (50%) of them being interval cancers, means that the remaining 100% are either true cancers (50%) or overdiagnosed cancers (50%)). Other researchers have supported our findings in studies from Catalonia, Spain[28] and from New South Wales, Australia,[29] and shown that there may be even higher levels of overdiagnosis in France.[30]

For Denmark, our estimate of overdiagnosis was 33%.[27] Likely explanations for the lower estimate are that participation rates in Copenhagen have been well below the recommended 70%[93,145], and a deliberately conservative attitude towards micro-calcifications and recalls.[132] Contrary, opportunistic screening in the "control" population was uncommon and therefore cannot explain the difference.[146] The fact that the incidence of carcinoma *in situ* increased only slightly in the non-screened areas following the introduction of screening[27] supports that opportunistic screening is infrequent. Such lesions are rarely symptomatic and therefore vastly more common in screened than non-screened women. In 2007, *in situ* cancers constituted 21% of screen-detected cases in the UK in the age group 50-70 years.[74] For comparison, *in situ* cases constituted 2.2% of diagnoses in the non-screened areas of Denmark in 2003, our last year of observation.[27]

We corrected our estimate of overdiagnsis in Denmark for a decline in breast cancer incidence in women over 70 years, which reduced our estimate from 40% to 33%.[27] However, this decline was only present in Funen, whereas the incidence in older women was increasing in the same time period in both Copenhagen and the non-screened areas. The decline in Funen was small in absolute numbers and may have been due to chance. Copenhagen had a longer screening period, so a decline should first appear there. But participation was higher in Funen, which speaks for a true compensatory decline. Further follow-up will reveal if the decline persists.

Comparing breast cancer incidence in screened and non-screened areas requires that the two populations are similar, for example in terms of socio-economic status. However, as we looked at changes in trends over time and compensated for pre-screening incidence differences, it is more important whether there were substantial changes over the observation period than if there were differences *per se* at an individual time point. In Denmark, there are differences in socio-economic status and educational level, with high education and urbanicity predicting high incidence, but also high survival.[147] But comparing larger geographical regions, as we did, will dilute such differences. In any case, statistical correction for confounders is not without problems. For socio-economic differences, the choice of measure (e.g. income or educational level) is important and could influence results. Also, models assume a linear correlation between e.g. income and life-expectancy, although this is unlikely.[148]

The abrupt changes in breast cancer incidence coincide with the introduction of breast screening in both Copenhagen and Funen, and in any other country where this has been studied, at time points varying by more than a decade. This strongly suggests that these changes are caused by screening.

How overdiagnosis helps promote screening

Overdiagnosis causes what has been termed the "popularity paradox".[1] Women who experience overdiagnosis and overtreatment will believe that screening saved them, even though they have in fact been harmed. Thus, the more overdiagnosis, the more popular the programme will become.

In actual fact, the chance that an individual woman diagnosed through screening have had her life saved by that screen is between 3 and 13%, if the effect of screening is a 5% to a 25% reduction.[149,150] Invitations and scientific publications often point out the alarming rate at which the incidence of breast cancer is rising in Western countries and how this underlines the importance of screening, without noting that a large part of the increase is caused by

screening itself.[16] It has recently been summed up how the rising incidence and unchanged mortality rates observed for a number of cancers show that we are simply diagnosing more cancer without affecting prognosis, and without diagnosing lethal cancer earlier.[8]

Advancing the time of diagnosis can be unfortunate

A type of overdiagnosis that has received less attention is when screening advances the time of diagnosis without affecting the prognosis or treatment. Screening then causes months or years of lifetime to be converted from living in "ignorant bliss" into being a cancer patient. This is an unavoidable consequence of screening, but the extent of this type of overdiagnosis remains to be studied. It is certain to diminish the quality of life of those who experience it.

Implications for the use of mastectomies

In November 2011, a letter was published in the Lancet by 41 signatories, all of whom have some relation to breast screening.[151] It was a reaction to the recent criticism of breast screening and focused specifically on the research from the Nordic Cochrane Centre. As we noted in our reply[152], the only factual information in the letter was a statement that 27% of women with screen-detected cancers have a mastectomy, compared to 53% with clinically detected cancers. This may be true, but it is misleading and regrettably an often used type of comparison in breast screening.

The problem is that this compares apples with oranges. The calculation is made within a population where all are offered screening, which means that there is no "control group". The compliant women are simply compared to the non-compliant ones and those with interval cancers. But screen-detected breast cancers are different from those detected because of symptoms, for several reasons:

1. Attendees are preferentially those with an already favourable prognosis.[1] Screening "selects" those that would turn up quickly with symptoms of small cancers in the absence of screening, whereas those that would wait and present with larger cancers do not turn up.

2. Screening preferentially detects slow-growing lesions because there is more time to detect them in (length bias). Thus, screening "selects" small cancers, while the aggressive ones grow fast and "slip through the screen" to appear between rounds as large cancers.

3. Many of the small, screen-detected cancers are overdiagnosed. This inflates the number of breast conserving surgical interventions in the screened group, which "artificially" reduces the percentage of mastectomies.

We need to compare a population offered screening with one that is not and compare the rate of mastectomies, say, per 1,000 women in each population. We have three main sources for this.

1. The randomised trials. This is the most reliable source, and there were 20% more mastectomies in the screened group.[45]

2. Population-based data from Denmark. Again, more mastectomies were performed in the screened areas, and we published these data in 2011 Fig. 9).[18]

3. Population based data from Norway, which also had a gradual introduction of mammography screening. The same picture emerged as in Denmark, which we also published in 2011.[31]

In population-based studies, there may be geographical variation in the choice of treatment apart from that resulting from screening, but the consistency of the findings, also with those from the randomised trials, indicate that screening does in fact increase the use of mastectomies. The increase is due to overdiagnosis of invasive and *in situ* lesions, and the fact that screening does not reduce the occurrence of large invasive cancers.[21]

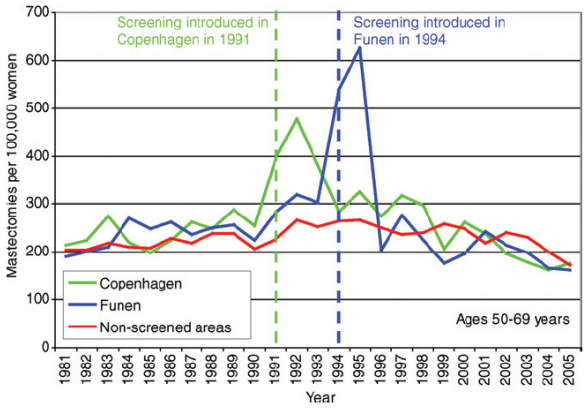

Figure 9. Mastectomy use in Copenhagen and Funen (screened areas), and non-screened areas. From Jørgensen et al 2011.[18]

Implications for radiotherapy use and mortality

In the randomised trials, the use of radiotherapy was increased to a similar extent as breast cancer incidence in the screened group, with a relative risk of 1.32 (95% confidence interval: 1.16 to 1.50).[45] The considerable overtreatment of healthy women with radiotherapy is a consequence of overdiagnosis. Radiotherapy not only causes long-term postoperative pain and skin-irritation[153], but also raises overall mortality through increased cardiac and lung-disease related mortality.[75,154,155] Radiation therapy can damage tissues and vessels and the harmful effects have recently been quantified for Denmark and Sweden for the period 1976-2006.[156] The study compared left- and right-sided radiotherapy. The relative risk of acute myocardial infarction was 1.22 (95% CI 1.06 to 1.42), angina 1.25 (1.05 to 1.49), pericarditis 1.61 (1.06 to 2.43), and valvular heart disease 1.54 (1.11 to 2.13). Women with previous heart disease had particularly high risks. Incidence ratios for all heart disease were as high for women irradiated since 1990 (1.09 [1.00 to 1.19]), as for women irradiated during 1976 to 1989 (1.08 [0.99 to 1.17]), indicating that modern radiation techniques may not have reduced harms as much as hoped for. Contrary to previous studies[75,154,155], the Scandinavian data did not indicate increased cardiac mortality, but the authors argue that this may be due to shorter follow-up, as clear effects on this outcome were only present after 15 years in other data sets that they have analysed.[154,155] Radiotherapy also considerably increase mortality from lung-disease, e.g. the relative risk was 2.71 (1.65-4.48) in North America.[155]

INFORMED CHOICE

Intentions and legislation

There is agreement that participation in screening should be voluntary and based on informed choice. This objective is specified in guidelines[145,157] and required by law in more general terms.[158,159] However, exactly what constitutes informed choice is subject to debate.[19,160] Is it voluntary to accept a pre-specified time for an investigation in an invitation from a public authority that directly recommends participation? This is a commonly used strategy[16] that is known to boost participation.[161] High participation is pivotal and it is not surprising that those who design the invitations choose this strategy, as they are often also responsible for the success of the programme. We have argued that this short-circuits informed decision[16,19] because a pre-specified appointment and a direct recommendation issued from a public authority will make some feel obliged to participate - a concern that now seems confirmed.[19,160]

The UK National Screening Committee agreed in 2001 that the purpose of invitations to screening was to ensure informed choice rather than high uptake.[162] However, the UK invitation and information leaflet still directly encourage participation, provides a pre-specified time of appointment, and are biased in favour of participation.[17,18] Encouraging participation was in line with a prior emphasis on uptake, as specified in the Forrest report:

"*Women up to age 65 years should be positively encouraged to be regularly screened (…)*"[13]

But such an approach conflicts with autonomy, which was clearly specified as a right by the General Medical Council.[157] It also creates problems when the image of simple and trouble-free screening collides with harsh reality. People who experience the harms feel let down and over-enthusiasm about screening can also make rational decisions about future screening programmes difficult[1].

Selling screening

There is no doubt that screening has been oversold in the past[16] (Fig. 10), but it is still oversold today.[18,33] Although the rhetoric seems more blunt in the United States than in Europe, the message in invitations are fundamentally identical.[16,17,18,163,164]

In 2007, a health reform in Germany for several cancer screening programmes, including breast screening, proposed that those who do not attend screening must attend mandatory personal counselling. If they still decline participation and are later diagnosed with the disease in question, they must pay 2% of their income towards treatment, compared to 1% for those who attend, if the invitees cannot document that they have participated in the mandatory counselling by presenting a signed form.[165,166] One can only wonder what information would have been offered during such counselling, but the bill was eventually turned down.

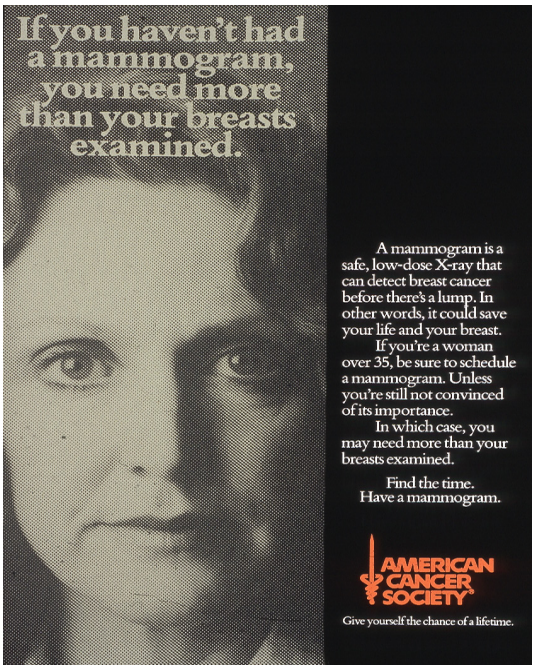

Figure 10: Campaign poster from the American Cancer Society.

Securing high uptake and informed choice at the same time is a difficult balance when there are both important benefits and harms. Screening proponents argue that it is necessary to provide a strong incentive to overcome what is commonly called "barriers to participation". Thorough information about harms is believed to deter some women. As invitations to screening are generally prepared by those responsible for the programme[16], this view has largely dictated their content. We have argued that being responsible for the screening programme is a conflict of interest and that invitations should be prepared by an independent committee.[16,17,19] Following our article in the spring of 2009 about the UK invitation and leaflet[17], a group of researchers from various fields wrote an open letter to The Times[167] pleading for improvement. Shortly after, the decision to revise the UK leaflet was announced.[168] The new leaflet was published around

Christmas 2010. Like the previous leaflet, the NHS BSP produced it themselves. We criticised the new leaflet, and the 2010 Annual Review of the NHS Breast Screening Programme,[18] and it has now been announced that the new leaflet will also be revised.[36]

Given the debate over the magnitude of benefits and harms, it is unsurprising that it is so difficult to reach consensus about what information should be provided. However, there are obvious shortcomings in the current invitations and information material. It is unreasonable to begin the invitation by scaring potential participants with the number of breast cancers and deaths per year. This information is meaningless if it is not related to the population size, and it does not make it clear that screening can affect only some of these deaths, as only a limited age range is invited and many cancers are detected between screening rounds. This information is often followed by a promise that screening can reduce breast cancer mortality by one third.[16] But this refers to the most optimistic trials and will make some invitees believe that screening can cut the total number of breast cancer deaths, which they just read about, by one third.

There is good evidence that absolute numbers rather than a relative risk will make people less likely to overestimate an effect.[169] As most information material present the benefits in statistical terms that are largely unintelligible to people other than statisticians with a sound knowledge of the background risks[1], it is important to provide invitees with absolute risk reductions and the absolute number of people who will experience a certain outcome, beneficial and harmful alike. We have argued that it is necessary to use the same denominator for both benefits and harms to make them comparable.[16] We have published a leaflet with information presented in this way, saying that 2,000 women must be screened over ten years to benefit one, whereas screening at the same time leads to ten women being overdiagnosed.[17] The leaflet is now available in 13 languages because people from various countries found it useful and offered to translate it (see: www.cochrane.dk). An updated version including the newest evidence was published in January 2012 in Danish and English.

Deciding what information to include and how to present it contains an element of subjectivity, although we do have some empirical evidence to support our choices, e.g. to use absolute rather than relative risks. This element of subjectivity is also important to keep in mind when evaluating the content of invitations and other information material and requires that this is done by two independent authors in a blinded fashion, which is how we did it when evaluating both web sites[15] and invitations.[16] An example from the latest invitation in the UK is its description of overdiagnosis:

"*Screening can find cancers which are treated but which may not otherwise have been found during your lifetime.*"[18]

The word "overdiagnosis" is not mentioned, nor is the magnitude of the problem. We have argued that this can be misleading, as some women might think "great, screening finds cancers that otherwise would not be found. That is why I go for screening".[18]

Providing balanced information is not only important for informed choice. It is important for continued trust in the medical profession.[1,170] Informing openly about the consequences of screening will prevent disappointment when the information about the harms becomes more widely known. And the concern that such information should prevent participation is counter to the available evidence from randomised trials of decision aids in breast and diabetes screening.[169,171]

A truly informed choice is not realistic for all eligible women. The issue is highly complex, some things are counterintuitive, and developing information that would be understandable to everyone may not be possible. But this does not change the objective, and access to unbiased and balanced information is more important than the (unsupported) risk of lower uptake.

CONFLICTS OF INTEREST IN MAMMOGRAPHY SCREENING

Evaluations of medical interventions are influenced by conflicts of interest. These can be economical, political, or driven by personal motives or beliefs. All are at play in mammography screening. Great economical interests are involved in a programme that rests on popular political decisions and the implementation of these decisions form the basis for medical careers, sought by people who believe in the rationale behind the intervention.[1]

The economic incentives are very large indeed. The UK screening programme is currently introducing digital mammography units at a cost of 100 million pounds[74] and in the United States, the breast screening industry turns over more than 5 billion US Dollars each year.[2] But the cost of the equipment and staff is only a part of the total cost. Diagnostic follow up tests, the treatment of overdiagnosed cases, and absence from work carry great financial costs, in addition to physical and psychological costs.

The industry has many ways to influence the introduction and expansion of screening programmes, some of which were summed up by Consultant for the UK National Screening Programmes Angela Raffle and former Programmes Director of the UK National Screening Committee, Muir Gray:

"[The industry] uses all kinds of strategies to promote sales of medical products. These include funding of conferences and scientific meetings, provision of ghost writers to get positive findings published quickly, suppression of publication of negative findings, direct lobbying of government, funding of patient pressure groups, assisting patient advocates with appeals against policy decisions and use of public relations firms to engineer a steady trickle of good news stories featuring individual cases claiming to have been helped by the technology. In the trade, the technique that involve pressure groups and patient advocates are known as 'astrosurfing' and 'guerilla marketing' because they successfully create the impression that the lobbying is coming purely from grassroots opinion and not from the industry."[1]

Although marketing from the screening industry is less recognised as an important problem than when pharmaceutical companies are involved, it is an obvious source of bias and anyone who attends a screening conference will notice it.

But there are conflicts of interest that are more subtle and harder to recognise. Those involved in the screening programmes harbour an inherent conflict of interest as their career hinges on it. Maureen Roberts was a lead researcher in the UK evaluation of breast cancer screening and clinical director of the Edinburgh Breast Screening Project. She died of breast cancer in 1989 but published an article in BMJ shortly before. She formulated what Angela Raffle and Muir Gray describe as something that, *"many of us involved in screening at that time will recognize as accurate."*[1]:

"There is an air of evangelism, few people questioning what is actually being done. Are we brainwashing ourselves into thinking that we are making a dramatic impact on a serious disease before we brainwash the public? Many thousands of women will be invited for screening and those who attend are said to be 'compliant'. The compliance rate is not very high and I wonder what plans are being made to try and raise it. I hope very much that pressure is not put on women to attend. The decision must be theirs, and a truthful account of the facts must be made available to the public and the individual patient. It will not be what they want to hear."[172]

Screening marketed as a product

The glorification of screening is clearly apparent in the 2008 Annual Report from the NHS BSP.[74] The report is issued from a public authority and is not directly influenced by commercial interests. The editor is Julietta Patnick, Director of the NHS BSP.

The overall message was that mammography screening has been hugely successful throughout its 20 years in existence, that benefits far outweighs harms (the most important of which are left entirely unmentioned), and participation is directly recommended. The format of the report would pride any marketing department, featuring a pink rose held up against a bright, revealing light, such as that used when scrutinizing mammograms (Fig. 11). The symbolism is strong; the rose for the female element; pink symbolising the fight against breast cancer; the bright light illustrating the knowledge and enlightenment through which the programme operates. Crucial critical questions were not answered, and later Annual Reports suffer from the same problems.[18]

Figure 11. Cover page from the 2008 Annual Review of the NHS Breast Screening Programme.[74]

The headline on the title page of the 2008 report reads: "Saving lives through screening". As described earlier in the section on breast cancer mortality, there is no evidence from the randomised trials that mammography screening saves lives in absolute terms. The providers optimistically assume that a reduction in breast cancer mortality translates directly into saved lives and they also disregard that overdiagnosis may cause deaths through overtreatment. This is a symptomatic simplification. There are numerous similar examples in the report, but its most important shortcoming is that it mentions none of the important harms: overdiagnosis and the psychological harms experienced by the many with false positive mammograms. Raffle and Gray note that:

"Governments are elected to deliver what people want. People believe that screening will harmlessly eliminate disease and pay for itself by reducing the need for treatment. The wise politician, who only thinks as far as the next election (or the next week in a ministerial post), inevitably makes decisions that match what the people believe in. Our challenge, and it is a considerable one, is to communicate and advocate public health evidence and values through the mass media so that public opinions come closer to our evidence-based view of the world."[1]

Politicians may very well introduce and maintain interventions that are contrary to the best available evidence, but in accordance with public opinion. To the public, screening enables an active effort against a feared disease. Being able to do something actively against a threat is a strong human motivator and aggressive action sometimes seems a guiding principle, particularly in North American medicine.

Is breast screening "worth it"?

Obviously, many of those saved from a breast cancer death by screening will eventually require treatment for another disease. Assuming that screening saves money because early detection means less treatment and less time at the hospital is therefore a simplification. Those who live on to collect their pensions will cost more than those who left society earlier. This, of course, is not an argument against screening, but it highlights the fallacy of equalling lower breast cancer mortality now with savings in the future. In addition to this, false positive cases and overdiagnosed cancers are certain to increase public spending considerably in the short term. These factors, and overdiagnosis in particular, have been underestimated in cost-benefit analyses[13,173-176] and it is unlikely that most screening programmes will reduce costs in the long run.[1] The problem is that screening proponents have effectively convinced the public and politicians that screening does save both lives and money, which can make arguments against screening seem like nitpicking.

Of course, if breast screening does not reduce breast cancer mortality or mastectomies, as current evidence suggests, it is meaningless to talk about health economics. In any case, it is very difficult to sum up the economic impact. A recent study used data on benefits and harms from the Cochrane review[45] and the U.S. Preventive Services Task Force Review[173] to calculate the impact of the programme on Quality Adjusted Life Years (QUALYs).[177] Harms outweighed benefits up to 7 years after screening implementation. Extrapolating the trial results to 20 years after implementation, which is not without concerns, the authors showed a net benefit, but less than half of what was expected at the time of the introduction of screening. But how do you quantify the human cost of being needlessly diagnosed with and treated for breast cancer? The authors set this to a 6% reduction in QUALYs, but such human costs require assumptions. Furthermore, the analysis built on an assumed 15% reduction in breast cancer mortality, which is unlikely to be correct in today's setting.

Quantifying bias in scientific papers

Although it is specified in the CONSORT guidelines that randomised trials of medical interventions should present evidence on both benefits and harms[178], this is often not done and many articles promote a specific agenda.[179-81] The imbalance in the presentation of benefits and harms and the relation to potential conflicts of interest has been well documented for both medical interventions and tobacco.[182,183] It has also been documented that industry sponsoring of trials of breast cancer interventions affects the study design and leads to more positive findings.[184]

It is hardly surprising that the mechanisms documented in other fields also operate in mammography screening. The main focus has been on the influence of the sponsor of the trial who has a direct financial interest in results. But industry funding of scientific papers on mammography screening is rare.

Our focus was on the involvement of authors associated with screening programmes, which are often publicly funded in Europe. We found that scientific articles tend to emphasize the benefits of mammography screening over its harms. This imbalance was related to the authors' affiliation, in particular for overdiagnosis, which was rarely mentioned in articles by authors working with screening compared to authors in other fields.[34] Emphasising benefits and neglecting harms was not limited to authors working with screening, but they downplayed or even rejected them particularly often. Likewise, the benefits were often presented using the most favourable framing, e.g. relative risks instead of absolute risks.[34]

It may seem surprising that being involved with screening can lead to bias, but clinical experts in a medical field are often biased. They may be particularly enthusiastic about an intervention, eager to improve outcomes for their own patients, or have affiliations with industry. In this case, questioning the screening programme is essentially equivalent to questioning their clinical specialty and livelihood. Funding must be secured for both the programme and their research, which may be provided by interest groups (e.g. cancer charities) that can have a preference for research that supports their political agenda. A mutually beneficial relationship can develop where positive results will be rewarded with more funding. This is particularly problematic if the interest group receives subsidies from the screening industry.

One example is The American Cancer Society (ACS) which is large enough to not only sponsor research, but also publish three scientific journals. We did a subgroup analysis of articles on mammography screening in the journal Cancer, one of those owned by the ACS. The subgroup analysis was not included in our published article[34], but it showed that articles in the journal Cancer were very positive towards breast screening; none of the articles raised criticism of mammography screening or acknowledged overdiagnosis as an important problem.

The ACS accepts industry funding from both the medical industry (AstraZeneca, Pfizer, Novartis) and health insurers:

"A new era of corporate outreach for the American Cancer Society has been launched through its Employer Initiative. Its goal is to build lasting relationships with major U.S. companies by offering and implementing products and services that help employers meet their business goals while increasing mission and income returns to the Society."[185]

There is an intricate web of co-operation and interdependence in the scientific community. Close collaborators are unlikely to criticise each other and personal relations can be hard to see through. The Journal of Medical Screening is a part of the Royal Society of Medicine Press, but it is also the journal of the Medical Screening Society. The Medical Screening Society is housed at the Wolfson Institute of Preventive Medicine, which is thus also the home institute of the Journal of Medical Screening. The Medical Screening Society has several members that were key figures in the implementation of screening.[186] The Wolfson Institute of Preventive Medicine is also the home institution of several of these screening proponents, some of whom also serve on the Editorial Committee of the Journal of Medical Screening. The Editorial Board also features the Director of the NHS Breast Screening Programme, Julietta Patnick. Professor Sir Nicholas Wald is both President of the Medical Screening Society, Director of the Wolfson Institute of Preventive Medicine, and Editor of Journal of Medical Screening. He thus heads the department where many of the screening proponents work who frequently publish in "his" journal.

We have criticised this hidden interdependence present in articles where the conclusions can be used politically to promote screening.[187,188] But such conflicts of interest are not generally recognised and are not mentioned in the articles in question, although the Journal of Medical Screening requires that conflicts of interest are disclosed. They are

therefore unlikely to be included in the considerations of the validity of the findings by readers who have no reason to expect such an intricate network.

On submission of a paper to BMJ, you are requested to disclose anything as conflicts of interest that you would not be happy to see publicised later in a wider context. This seems a good guiding principle.

New independent review in the UK

As in other fields of research, the independence of the researcher is paramount. Being your own judge is problematic, which is why police, legislators, and judicial authorities are kept separate. It must be required that independent experts critically evaluate publicly funded screening programmes. The importance of this was re-affirmed when we evaluated the 2010 Annual Review of the NHS Breast Screening Programme[18] and showed how skewed the official presentation still is.

Our research has influenced the decision to form an independent panel to review the evidence and continued justification for breast screening.[35,36] While the necessity of independent experts were recognised, these were appointed by Professor Sir Michael Richards, who, as Cancer Director in the NHS, is formally responsible for the breast screening programme, together with Harpal Kumar, the Chief Executive of Cancer Research UK, a cancer charity that is actively promoting breast screening. Independence is in this case defined as researchers who have not previously published in the field, which can be questioned as it does not preclude having preconceived opinions. We[189] and many others[190-94] questioned if the planned review would be truly independent, but the experts that were eventually chosen[195] indicate that the stated wish to have an independent review was honest rather than window dressing aimed at protecting the programme against criticism. The results of the review should appear in 2012.

ETHICAL CONSIDERATIONS

Some questions are not accessible to scientific enquiry, such as balancing harms against benefits, which is a value judgement. Ethical theory may help us clarify what principles our decisions are based on and help us assess if we are being consistent in our choices. It should be acknowledged, however, that there is no "correct" answer. People have different values and life experiences that will inevitably influence their judgment, and one decision is not necessarily better than another. In health care, anyone has the right to choose for him- or herself, and health authorities are obliged to provide information to secure autonomous choice regarding the interventions they decide to offer. These issues are of particular importance in screening, as this is intended for healthy individuals, not patients who actively seek the opinion of health professionals. Those invited have not requested an intervention to solve a specific problem and are not in a situation where an intervention is necessary to overcome disease.

Even though preventing breast cancer fatalities is an important good, the adverse effects of screening may be harmful enough to outweigh it. If we made prophylactic mastectomy mandatory for 40-year-old women, we would likely prevent most breast cancer fatalities. We have chosen not to do this because we find it obvious that the harms would be too great, without even bothering to express this explicitly.

This example is not as extreme as it seems, as screening mammography does indeed cause unnecessary removal of many breasts. Less extreme harms than prophylactic mastectomy could lead to a similar conclusion; that the preventive measure is not worth the human or financial cost. We have to be open to this possibility and constantly re-evaluate if the balance favours the intervention. Such a re-evaluation may soon become relevant in the case of cervical cancer screening where vaccination could reduce the obtainable benefit to an extent where the harms will outweigh it, in this case the many unnecessary conisations for self-limiting cell changes, which increase the risk of premature labour. However, it may also be relevant when new harms are recognised or better quantified, as for overdiagnosis in mammography screening.

Weighing the harms against the benefits in mammography screening is currently a concern for the invited woman. Evaluating the programme against other health interventions pose a problem that is at least as complex, but this is a concern for society. This question is becoming increasingly relevant with tightening budgets in health care. The decision to implement screening was made before overdiagnosis was recognised as a major harm and it was therefore not considered to a sufficient extent in the decision process or the cost-benefit analyses.[13] Since then, expectations of the obtainable benefit have been substantially reduced.[6,37,45,173] When the decision was made to implement screening, the Two-County trial carried great weight.[13,60,65] However, the reliability of this trial has been questioned and later trials of higher quality estimated much smaller effects.[45,66] The likely benefit is, at best, only half of what was hoped for. It may even no longer exist due to improved treatment[23,24] and the harms caused by treating overdiagnosed women with chemo- and radiotherapy.

FUTURE ASPECTS

New diagnostic technologies

Technological developments mean that mammography can detect ever smaller abnormalities. This is augmented by digital mammography, computer-aided detection, and MRI scans. It is easy to speculate that this will improve the ability of screening to prevent breast cancer fatalities since cancer can be caught earlier than before. However, the earliest trials of mammography screening (the New York Health Insurance Plan Trial and the Two-County trial) presented the most optimistic results, with later trials showing a smaller or no effect. While methodological problems in the first trials may explain some of the large effect, the absence of an effect in population-based screening programmes could also be due to new treatments such as anti-oestrogens (e.g. tamoxifen) that were not in use during the first trials, but were introduced during the 1980's.[40] Current research suggests that digital mammography is unlikely to offer an improvement over standard film, at least in the age group commonly invited for breast screening.[196] However, the ability to detect smaller abnormalities will increase the number of harmless lesions that need further work-up, and the number of overdiagnosed cases. In short, we can be fairly sure that there will be more harms, but not that the benefit will increase at a similar rate.

Hopefully, we will be better equipped to tell the potentially fatal lesions from the harmless ones in the future. So far, histology has proved too rough a tool to uncover what is likely genetic differences that determine the rate of progression and the metastatic potential of individual cancers. We may speculate that identification of genetic markers will provide a means to differentiate between the aggressiveness of individual cancers and offer more individualised treatment, especially as mammographically detected tumours now seems to be distinguishable as having a favourable genetic profile.[197] However, the value of such tests in terms of better outcomes for the patient is still unproven and their implementation is therefore only a theoretical possibility so far. Even if we could predict the course of each individual lesion down to the point where we could inform a patient that "your screen-detected tumour will become fatal in about 10-14 years", it would still not make overdiagnosis disappear, only reduce the problem. In fact, most women would likely choose to be treated anyhow when the cancer has been detected.

Breast cancer screening in high-risk groups

In the light of the recognition that mammography screening offers smaller benefits and are more harmful than we were promised, it has been suggested that screening should be limited to groups with a high risk of breast cancer.[198] The idea is that the balance between benefits and harms may "tip" in a favourable direction because the higher frequency of disease will increase the chance of a benefit. But this requires the assumption of an unchanged relative chance of benefit and harm in this subgroup.

Using BRCA mutation carriers as an example, we can see that things, as usual, are a bit more complicated. The mutation is in a gene that codes for a DNA repair mechanism, which substantially increases the risk that genetic mutations accumulate uncorrected and leads to cancer. But mammography screening at short intervals to increase the chance of catching these particularly fast-growing and aggressive cancers will also subject these women to more radiation. While this is considered a small problem in the general population, it is much worse in BRCA-carriers. A recent study calculated that in women aged 25-29 years, screening would have to cut breast cancer mortality in half to outweigh the increased risk induced by the X-rays.[199] While this is below the usual screening age, and despite that the required reduction was much smaller in older women, such young women with BRCA mutations (or other markers of high risk) are those most relevant to target as they generally develop aggressive breast cancer at a very early age. Young women are inherently more sensitive to radiation because of rapid cell turn-over. They also have denser breast tissue, which not only makes the mammograms more difficult to interpret, leading to less benefit, but also increases the risk of false positives and possibly overdiagnosis.

We should therefore require randomised trials before decisions about screening high-risk groups are made.

Screening for other cancers

Upcoming cancer screening programmes in Western societies are those for colorectal cancer, lung cancer with spiral computed tomography, and prostate cancer, and a trial of lung cancer screening are currently being carried out. The issues dealt with in this thesis are also relevant to these screening programmes, although each must be evaluated in their own right.

Regarding screening for prostate cancer, there are marked differences in the approach in North America and Scandinavia. In North America, screening for prostate cancer with the prostate specific antigen (PSA) test is common, whereas the Urological Society in Denmark recommends against it.[200] Previous randomised trials of PSA screening are of poor quality[201] and new, well-designed trials show low effects on mortality and much overdiagnosis, with important harms related to treatment such as incontinence and impotence.[113,114] New U.S. Preventive Services Task Force Draft Guidelines recommend against it.[115]

While we are still awaiting evidence from several on-going, large-scale randomised trials of lung cancer screening with low-dose computed tomography (LDCT), a recently published randomised trial of 2,472 male smokers concluded that the benefits might be far smaller than anticipated.[202] All subjects received a baseline chest X-ray and sputum cytology test, but the intervention arm then went through LDCT every year for the next four years. There were no differences in lung cancer mortality or total mortality after a median follow-up of 33 months, but lung cancer was diagnosed in 4.7% in the intervention arm versus 2.8% in the control arm. Additional cancers diagnosed in the screened arm were stage I disease, with no reduction in advanced stage lung cancer. It requires further follow-up to determine whether the extra cancers were overdiagnosed. But if the 70% increase in lung cancer incidence were overdiagnosis, this would be the same level as in prostate cancer screening with PSA.[113] Observational studies have reached very positive conclusions about LDCT screening[81], which illustrates the need for randomised trials.

Randomised trials have shown a benefit from screening for colorectal cancer with faecal occult blood tests[203] and also with once-only sigmoidoscopy.[49] Although overdiagnosis may be a small problem related to the detection of benign polyps, false positive and negative cases remain a problem and there are significant, albeit rare, harms caused by sigmoidoscopy and follow-up interventions, such as perforations, bleeding, and thrombosis associated with sedation.

Private clinics, also in Denmark[204], offer screening for breast cancer using thermo-mammography and other techniques as a supplement to regular screening mammography without evidence for a benefit and certainty of harm. We are likely to see more such interventions in the future intended for the "wealthy worried well".[1]

HISTORICAL ASPECTS

Mammography screening has been called "a crisis for evidence based medicine"[205] because its immediate appeal may cause the evidence to be improperly interpreted.

It can be quite a learning experience to re-visit previous beliefs. The Forrest Report that laid the foundation for the UK Breast Screening Programme mentions overdiagnosis in several places, e.g.:

"…*women might undergo unnecessary procedures for the diagnosis and treatment of cancer which might not have entered an invasive phase during their lifetime.*" [13]

But the report considers it unlikely to be a great problem based on experience from the New York Health Insurance Plan trial, which showed that the same amount of cancer was detected in its two arms, likely because more women with a breast cancer diagnosis prior to screening were excluded from the intervention-arm. The report acknowledge that 20% more cancers were detected in the intervention arm of the only other trial available at the time, the Two-County trial, but explained this with too short follow-up. The report goes on:

"*If screening were detecting breast cancers that would otherwise not have been diagnosed, it would be expected that in controlled trials there would be a persistent excess number of breast cancers in the screened group compared with the control group.*"[13]

This is absolutely true, and is how overdiagnosis was quantified in the recent prostate cancer screening trials, without much debate.[113,114] But when it became clear that there was indeed a persistent excess incidence in the trials of mammography screening published after the New York trial, as well as in public screening programmes, the explanation was changed so that the excess incidence was considered early diagnosis to be compensated later by a drop in incidence in previously screened women (Fig. 7).[139] We have demonstrated that such a compensation does not occur to any substantial extent, even in the UK where screening has operated for more than 20 years (Fig. 8).[18,26] The premises were changed as the facts that contradicted them became obvious.

We may also ask if it is reasonable that the New York Health Insurance Plan trial, which found no excess incidence of breast cancer, really advanced the time of diagnosis. It found one of the largest effects, but according to screening theory, a large effect requires a substantial advancement of diagnosis. No screening proponent has bothered to explain this contradiction. Extreme results often appear in the first studies and this may skew our perception of the true effect.[206] A preference for citing only the positive trial results is known as citation bias, or "optimism bias" in a slightly wider context that also include e.g. selective citing of positive results within the trials.[207] The authors suggested that systematic reviews are the best way to avoid this bias.

It has been claimed that the 30-50% higher incidence observed in populations with organised screening programmes was due exclusively to such an advancement of the time of diagnosis, but this is unreasonable based on assumptions of an average lead time of even 2.5 years. Women who are 52.5 years do not have an incidence that is 30-50% higher than women who are 50 years. When has enough time elapsed before we acknowledge that the increase in incidence is due to overdiagnosis?

CONCLUSIONS

The decision to implement public screening programmes for breast cancer with mammography was based on an overestimate of the benefit. The most important harm, overdiagnosis, was insufficiently recognised, and the premise that screening leads to less invasive treatment was wrong.

The benefit has been oversold to the public and the harms have been downplayed or neglected not only in information material, but also in scientific research. Invitations to screening have had the objective to increase uptake rather than to promote informed choice, but it is not acceptable to neglect the requirement for autonomy.

Overdiagnosis is not limited to *in situ* cancers; in fact there are more overdiagnosed invasive breast cancers. Although the level of overdiagnosis is still debated, it fundamentally changes the way mammography screening should be evaluated and challenges the justification for breast screening.

We must be honest about overdiagnosis, and about the reduced benefit compared to what we hoped for when screening mammography was introduced. This is the only way we can ensure informed choice. It is therefore necessary that information material and invitations are prepared by impartial entities and not by those offering screening. A clear message must be sent that screening may not reduce the risk of dying from breast cancer, that attendance considerably increases the risk of receiving a breast cancer diagnosis and a mastectomy, and that abstaining from screening can therefore be a sensible choice for many women.

It is necessary to constantly re-evaluate the merits and justification for medical interventions, in particular when the benefits are small and the harms are common and serious, and when new evidence challenges previous beliefs. To ensure that future evaluations of the continued justification for mammography screening are neutral, it is paramount that conflicts of interest are avoided. The upcoming revision of the invitation to breast screening in the UK and the announcement of an independent review of the evidence and justification for the intervention represents an important recognition of this.

REFERENCES

1. Raffle A, Gray M. *Screening: evidence and practice*. Oxford: Oxford University Press, 2007.
2. Editorial. More Questions About Mammograms. *The New York Times*, 26 October 2011 http://www.nytimes.com/2011/10/27/opinion/more-questions-about-mammograms.html?_r=1&scp=5&sq=editorial%2026%20october%202011&st=cse.
3. U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med* 2009;151:716-26.
4. Kolata G. Behind Cancer Guidelines, Quest for Data. *The New York Times*, 22 November 2009. http://www.nytimes.com/2009/11/23/health/23cancer.html?_r=1.
5. DeAngelis CD, Fontanarosa PB. US Preventive Services Task Force and Breast Cancer Screening. *JAMA* 2010;303:172-3.
6. The Canadian Task Force for Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ* 2011;DOI:10.1503/cmaj.110334
7. Raffe MJ, Gordon P, Rakovitch E, Kopans D. Screening recommendations: we must use science, relevant modern data, and common sense in balancing benefit and harm. Rapid response *CMAJ* 2011. http://www.cmaj.ca/content/183/17/1991.full#responses.
8. Esserman L, Shieh Y, Thompson I. Rethinking Screening for Breast Cancer and Prostate Cancer. *JAMA* 2009;302:1685-92.
9. McNeil C. Magnitude of Overdiagnosis in Cancer Indicate Need for Strategies to Address the Problem. *J Natl Cancer Inst* 2010;102:583.
10. Olsen AH, Jensen A, Njor S, Villadsen E, Schwartz W, Vejborg I, Lynge E. Breast cancer incidence after the start of mammography screening in Denmark. *Br J Cancer* 2003;88:362-65.
11. Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, Myles JP, Warwick J. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* 2005;7:258-65.
12. Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen THH, Yen AMF, Smith RA. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomised trial, and from the Breast Screening Programme in England. *J Med Screen* 2010;17:25-30.
13. Forrest P. *Breast Cancer Screening. Report to the Health Ministers of England, Wales, Scotland & Northern Ireland.* London: Department of Health and Social Science, 1986.
14. Vainio H, Bianchini F. IARC *Handbooks of Cancer Prevention Vol. 7: Breast Cancer Screening.* Lyon: IARC Press, 2002.
15. Jørgensen KJ, Gøtzsche PC. Presentation on web sites of possible benefits and harms from screening for breast cancer: cross sectional study. *BMJ* 2004;328:148-51.
16. Jørgensen KJ, Gøtzsche PC. Content of invitations for publicly funded screening mammography. *BMJ* 2006;338:538-541.
17. Gøtzsche PC, Hartling OJ, Nielsen M, Brodersen J, Jørgensen KJ. Breast screening: the facts – or maybe not? *BMJ* 2009;338:446-48.
18. Gøtzsche PC, Jørgensen KJ. The NHS Breast Screening Programme continues to misinform the public, ignoring repeated criticism and important new evidence. *J R Soc Med* 2011;104:361-9.
19. Jørgensen KJ, Brodersen J, Hartling OJ, Nielsen M, Gøtzsche PC. Informed choice requires information about both benefits and harms. *J Med Ethics* 2009;35:268-9.
20. Barry MJ, Mulley AJ. Why are high overdiagnosis probability and long lead time for prostate screening so important? *J Natl Cancer Inst* 2009;101:362-3.
21. Autier P, Boniol M, Middleton R, Doré J-F, Héry C, Zheng T, *et al.* Advanced breast cancer incidence following population based mammographic screening. *Ann Oncol* 2011. doi:10.1093/annonc/mdq633.

22. Gøtzsche PC, Jørgensen KJ, Zahl PH, Mæhlen J. Why mammography screening hasn't lived up to expectations from the randomised trials. *Cancer Causes Contr* 2011; DOI 10.1007/s10552-011-9867-8.
23. Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark. A comparative study. *BMJ* 2010;340:c1241.
24. Kalager M, Zelen M, Langmark F, Adami HO. Effect of Screening Mammography on Breast-Cancer Mortality in Norway. *N Engl J Med* 2010;363:1203-10.
25. Autier P, Boniol M, Gavin A, Vatten LJ. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. *BMJ* 2011;343:d4411 doi: 10.1136/bmj.d4411.
26. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
27. Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009;9:36.
28. Martinez-Alonso M, Vilaprinyo E, Marcos-Gragera R, Rue M. Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res* 2010;12(4):R58.
29. Morrell S, Barratt A, Irwing L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control* 2010;21:275-82.
30. Junod B, Zahl PH, Kaplan RM, Olsen J, Greenland S. An investigation of the Apparent Breast Cancer Epidemic in France: Screening and incidence trends in birth cohorts. *BMC Cancer* 2011;11:401.
31. Surhke P, Mæhlen J, Schlichting E, Jørgensen KJ, Gøtzsche PC, Zahl PH. Mammography screening and surgical breast cancer treatment in Norway: comparative analysis of cancer registry data. *BMJ* 2011; 343:d4692.
32. Jørgensen KJ, Keen J, Gøtzsche PC. Is Mammographic Screening Justifiable Considering its Substantial Overdiagnosis Rate and Minor Effect on Mortality? *Radiology* 2011;260:621-7.
33. Jørgensen KJ, Gøtzsche PC. Who evaluates public health programmes? The NHS Breast Screening Programme reviews the NHS Breast Screening Programme. *J R Soc Med* 2010;103:14-20.
34. Jørgensen KJ, Klahn A, Gøtzsche PC. Are benefits and harms given equal attention in scientific articles on mammography screening? A cross-sectional study. *BMC Medicine* 2007;5:12.
35. Bewley S. The NHS breast screening programme needs independent review. *BMJ* 2011; 343:d6894.
36. Richards M. An independent review is under way. *BMJ* 2011; 343:d6843.
37. Humphrey LL, Helfand M, Chen BK, Woolf SH. Breast cancer screening: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002;137:347-60.
38. Engel J, Eckel R, Kerr J et al. The process of metastasisation for breast cancer. *Eur J Cancer* 2003;39:1794–1806.
39. Baum M, Demichelli R, Hrushesky W, Retsky M. Does surgery unfavourably perturb the "natural history" of breast cancer by accelerating the appearance of distant metastases? *Eur J Cancer* 2005;41:508-15.
40. Baum M. *Breast Beating.* Kent: Anshan Ltd., 2010.
41. Reis-Filho JS, Pusztai L. Gene expression profiling in breast cancer: classification, prognostication, and prediction. *Lancet* 2011;378:1812-23.
42. Welch GH. *Should I be tested for cancer?* California: University of California Press, Ltd; 2004.
43. Haakinson DJ, Chee-Chee HS, Dueck AC, Gray RJ, Wasif N, Apsey HA, Pockaj BP. A significant number of women present with palpable breast cancer even with a normal mammogram within 1 year. *Am J Surg* 2010;200:712-18.

44. Bailey SL, Sigal BM, Plevritis SK. A simulation model investigating the impact of tumor volume doubling time and mammographic tumor detectability on screening outcomes in women aged 40-49 years. *J Nat Cancer Inst* 2010;102:1263-71.
45. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews* 2009, Issue 4. Art. No.:CD001877.
46. Folkman J, Watson K, Ingber D, Hanahan D. Induction of angiogenesis during the transition from hyperplasia to neoplasia. *Nature* 1989;339:58-61.
47. Sørlie T, Perou CM, Tibshirani R et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci* 2001;98:10869-74.
48. Narod S. Age of diagnosis, tumor size, and survival after breast cancer: implications for screening. *Breast Cancer Res Treat* 2011;128:259-66.
49. Atkinson WS, Edwards R, Kralj-Hans I, Wooldrage K, Hart AR, Northover JMA, Parkin DM, Wardle J, Duffy SW, Cuzick J. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010;375:1624-33.
50. Mukherjee S.*The Emperor of All Maladies*. New York: Fourth Estate; 2011.
51. Giuliano A, Hunt KK, Ballman KV et al. Axillary Dissection vs No axillary Dissection in Women With Invasive Breast Cancer and Sentinel Node Metastasis. *JAMA* 2011;305:569-75.
52. Breast Cancer Study Group (IBCSG). Trial 23-01. 34th Annual San Antonio Breast Cancer Symposium (SABCS): Abstract S3-1. Presented December 8, 2011.
53. Rettig RA, Jacobson PD, Farqhuar CM, Aubry W. *False Hope. Bone Marrow Transplantation for Breast Cancer*. New York: Oxford University Press, 2007.
54. Farquhar C, Marjoribanks J, Basser R, Lethaby A. High dose chemotherapy and autologous bone marrow or stem cell transplantation versus conventional chemotherapy for women with early poor prognosis breast cancer. *Cochrane Database of Systematic Reviews* 2005, Issue 3. Art. No.: CD003139. DOI: 10.1002/14651858.CD003139.pub2.
55. Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al.14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999;353:1903-8.
56. Bjurstam N, Björneld L, Warwick J, Sala E, Duffy SW, Nyström L. The Gothenburg Breast Screening Trial. *Cancer* 2003;97:2387-96.
57. Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ* 1988;297:943-8.
58. Shapiro S, Strax P, Venet L. Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA* 1971; 215: 1777-85.
59. Frisell J, Eklund G, Hellstrom L, Lidbrink E, Rutqvist LE, Somell A. Randomized study of mammography screening - preliminary report on mortality in the Stockholm trial. *Br Cancer Res Treat* 1991;18:49-56.
60. Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 1985;1:829-32.
61. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. *J Natl Cancer Inst* 2000;92:1490-9.
62. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002;137:305-12.
63. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L, for the Trial Management Group. Effect of mammographic screening

from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006;368:2053-60.

64. Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews* 2001, Issue 4. Art. No.: CD001877.
65. Statusrapport. [*Early detection and treatment of breast cancer*]. Sundhedsstyrelsen, Copenhagen 1997.
66. Zahl PH, Gøtzsche PC, Andersen JM, Mæhlen J. Results of the Two-County trial of mammography screening are not compatible with contemporaneous official Swedish breast cancer statistics. *Dan Med Bull* 2006;53:438-40.
67. Gøtzsche PC, Mæhlen J, Zahl PH. What is publication? *Lancet* 2006;**368:**1854-6.
68. Gøtzsche PC, Mæhlen J, Zahl PH. Undeclared motives in withdrawing a publication. *Lancet* 2007;369:1690.
69. Crewdson J. Swedes doubt mammography trial. Disparities found in landmark study. *Chicago Tribune*, 15 March 2002.
70. Holmberg L, Duffy SW, Yen AM, Tabár L, Vitak B, Nyström L, Frisell J. Differences in endpoints between the Swedish W-E (two-county) trial of mammographic screening and the Swedish overview: methodological consequences. *J Med Screen* 2009;16:73-80.
71. Domenighetti G, D'Avanzo B, Egger M, Berrino F, Perneger T, Mosconi P, et al. Women's perception of the benefits of mammography screening: population-based survey in four countries. *Int J Epidemiol* 2003;32:816-21.
72. Schwartz LM, Woloshin S, Fowler FJ Jr, Welch HG. Enthusiasm for cancer screening in the United States. *JAMA* 2004;291:71-8.
73. Gigerentzer G, Mata J, Frank R. Public knowledge of benefits of breast and prostate cancer screening in Europe. *J Natl Cancer Inst* 2009;101:1216-20.
74. Patnick J (ed.). *Breast Screening Programme. Annual Review 2008. Saving lives through screening.* Available at: http://www.cancerscreening.nhs.uk/breastscreen/publications/2008review.html. Accessed 17 May 2010.
75. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* 2005;365:1687–1717.
76. Blamey RW, Ellis IO, Pinder SE et al. Survival of breast cancer according to the Nottingham Prognostic Index in cases diagnosed in 1990-1999. *Eur J Cancer* 2007;43:1548-55.
77. Rostgaard K, Vaeth M, Rootze´n H, Lynge E. Why did the breast cancer lymph node status distribution improve in Denmark in the pre-mammography screening period of 1978–1994? *Acta Oncol* 2010;49:313–321.
78. Rose G. *Strategy of preventive medicine.* Oxford Universtity Press, 1992.
79. Badgwell BD, Giordano SH, Duan ZZ, et al: Mammography before diagnosis among women age 80 years and older with breast cancer. *J Clin Oncol* 2008;26:1–8.
80. Berry D, Baines CJ, Baum M, Dickersin K, Fletcher SW, Gøtzsche PC, Jørgensen KJ**,** Junod B, Mæhlen J, Schwartz LM, Welch GH, Woloshin SH, Thornton H, Zahl PH. Flawed Inferences About Screening Mammography's Benefit Based on Observational Data. *J Clin Oncol* 2009;27:639-40.
81. Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, Miettinen OS, International Early Lung Cancer Early Action Program Investigators. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med.* 2006;355:1763-71.
82. Welch HG, Woloshin SW, Schwartz LM, Gordis L, Gøtzsche PC, Harris R, Kramer BS, Ransohoff DF. Overstating the evidence for lung cancer screening. *Arch Intern Med* 2007;167:2289-95.
83. Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB et al. Breast cancer mortality in Copenhagen after introduction of mammography screening: Cohort study. *BMJ* 2005;330:220-4.

84. Gøtzsche PC, Thornton H, Jørgensen KJ. Reduction in mortality from breast cancer: Presentation of benefits and harms needs to be balanced. *BMJ* 2005;330:1024.
85. Janzon L, Andersson I (1991) The Malmo¨ mammographic screening trial. In: *Cancer Screening*, Miller AB, Chamberlain J, Day NE et al eds). pp 37–44. Cambridge University Press: Cambridge.
86. van Schoor G, Moss SM, Otten JDM, Donders R, Paap E, den Heeten GJ, Holland R, Broeders MJM, Verbeek ALM. Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* 2011;104: 910–914.
87. Jørgensen KJ. Flawed methods explain the effect of mammography screening in Nijmegen. *Br J Cancer* 2011; doi:10.1038/bjc.2011.264.
88. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA* 2000;283:2975–8.
89. Methling I, Bech-Jessen F. [Cancer screening saves lives] *Politiken*, 13 January 2005, pp 1.
90. Danish Breast Cancer Cooperative Group. [*Quality Indicator Repport for Breast Cancer 2008*]. Danish Breast Cancer Cooperative Group, Copenhagen 2009.
91. Pedersen FS. [Cancer victims to be compensated for unnecessary breast surgery]. *Berlingske Tidende* 25 January 2007.
92. Nyström L, Anderson I, Bjurstam N, Frisell J, Nordenskjold B, Rutquist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. *Lancet* 2002;359:909-19.
93. Euler-Chelpin M, Olsen AH, Njor S, Vejborg I, Schwartz W, Lynge E. Women's patterns of participation in mammography screening in Denmark. *Eur J Epidemiol* 2006;21:203-9.
94. Lynge E, Olsen AH, Njor S. Rapid response. *BMJ* 25 March 2010. http://www.bmj.com/cgi/eletters/340/mar23_1/c1241.
95. Olsen AH, Njor SH, Lynge E. Estimating the benefits of mammography screening. The impact of study design. *Epidemiology* 2007;18:487-92.
96. Autier P, Boniol M, LaVecchia C, Vatten L, Gavin A, Héry C, et al. Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ* 2010;341:c3620.
97. Black WC, Haggstrom DA, Welch GH. All-Cause Mortiality in randomised trials of cancer screening. *J Nat Cancer Inst* 2002;94:167-73.
98. Rossouw JE, Anderson GL, Prentice RL, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-333.
99. Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003; 362: 419-427.
100. Zahl PH, Mæhlen J. A decline in breast cancer incidence. *N Engl J Med* 2007;357:509-13
101. Ravdin PM, Cronin RA, Howlander N et al. The decrease in breast cancer incidence in 2003 in the United States. *N Engl J Med* 2007;356:1670-74.
102. Elfenbein GJ. A decline in breast cancer incidence. *N Engl J Med* 2007;357:509-13.
103. Thorat M. What caused the decline in US breast cancer incidence. *Nat Clin Pract Oncol* 2008;5:314-5.
104. Politiken. *Viden* pp. 2, 23 October 2011.
105. Parvinen I, Chiu S, Pylkkänen L et al. Effects of annual vs triennial mammography interval on breast cancer incidence and mortality in ages 40-49 in Finland. *Br J Cancer* 2011;105:1388-91.
106. Zahl PH, Mæhlen J, Welch HG. The natural history of breast cancers detected by screening mammography. *Arch Intern Med* 2008;168:2311-6.
107. Zahl PH, Gøtzsche PC, Mæhlen J. Natural history of breast cancer detected in the Swedish mammography screening programme: a cohort study. *Lancet Oncol* 2011 doi:10.1016/S1470-2045(11)70250-9.
108. Barry MD. Screening for Prostate Cancer – The controversy That Refuses to Die. *N Engl J Med* 2009;360:1351-4.

109. Draisma G, Etzioni R, Tsodikov A, et al. Lead time and overdiagnosis in prostate-specific antigen screening: importance of methods and context. *J Natl Cancer Inst* 2009;101:374–83.
110. Marcus PM, Bergstrahl EJ, Lung cancer mortality in the Mayo Lung Project: impact of extended follow-up. *J Natl Cancer Inst* 2000;92:1280-82.
111. Black WC. Overdiangosis: and underrecognized cause of confusion and harm in cancer screening. *J Natl Cancer Inst* 2000;92:1280-82.
112. Welch GH, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102:605-13.
113. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V et al. Screening and Prostate-Cancer Mortality in a Randomized European Sudy. *N Engl J Med* 2009;360:1320-8.
114. Andriole GL, Grubb RL, Buys SS, Chia D, Church TR, Fouad MN et al. Mortality Results from a Randomized Prostate-Cancer Screening Trial. *N Engl J Med* 2009;360:1310-9.
115. U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force Recommendation Statement. Available at: http://www.uspreventiveservicestaskforce.org/draftrec3.htm. Accessed 12 December 2011.
116. Welch GH, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? *Ann Intern Med* 1997;127:102-8.
117. Kaplan RM, Porzholt F. The natural history of breast cancer. *Arch Intern Med* 2008;168:2302-3.
118. Kolata G. Study Suggests Some Cancers May Go Away. *The New York Times*, 24 November 2008.
119. Clayton D, Hills M. *Statistical Models in Epidemiology*. Oxford University Press, Oxford 1993.
120. Larsen SU, Rose C. [Spontaneous remission of breast cancer: a literature review]. *Ugeskr Laeger*: 1999;161:4001-4.
121. Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittam V, Remington PL. The Wisconsin breast cancer simulation model. *J Natl Cancer Inst Monogr* 2006;36:37-47.
122. Joensuu H, Lundin J. Spontaneous regression of cancerous tumors detected by mammography screening. *JAMA* 2004:292;2580.
123. Schilling FH, Spix C, Berthold F, et al. Neuroblastoma screening at one year of age. *N Engl J Med* 2002;346:1047-53.
124. Kopans D. Why the critics of screening mammography are wrong. They distort data, rely on weak evidence, but refuse to defend when challenged. *Diagnostic Imaging* 2009;31:1-7.
125. Yen MF, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer* 2003;39:1746-54.
126. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42:793-806.
127. 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.
128. Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH. Complexities in the estimation of overdiagnosis in breast cancer screening. *Br J Cancer* 2008;99:176-8.
129. 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.
130. BBC Today: http://news.bbc.co.uk/today/hi/today/newsid_8596000/8596225.stm. 31 March 2010.
131. Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000; 355:129-34.
132. Svendsen AL, Olsen AH, von Euler-Chelpin M, Lynge E. Breast Cancer Incidence after the Introduction of Mammography Screening – What Should Be Expected? *Cancer* 2006;106:1883-90.
133. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer

overdetection with mammography screening: a systematic review. *Lancet Oncol* 2007;8:1129-38.
134. Zahl PH, Jørgensen KJ, Mæhlen J, Gøtzsche PC. Biases in estimates of overdetection due to mammography screening. *Lancet Oncol* 2008;9:199-201.
135. Gøtzsche PC, Mæhlen J, Jørgensen KJ, Zahl PH. Estimation of lead-time and overdiagnosis in breast cancer screening. *Br J Cancer* 2009;100;219.
136. 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.
137. Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F 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.
138. Zahl PH, Strand BH, Mæhlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004;328:921-4.
139. Boer R, Warmerdam P, de Koning H et al. Extra incidence caused by mammographic screening. *Lancet* 1994;343:979.
140. Botha JL, Bray F, Sankila R, Parkin DM. Breast cancer incidence and mortality trends in 16 European countries. *Eur J Cancer* 2003;39:1718-29.
141. Cancer Research UK. Breast cancer - UK incidence statistics. Available at: http://info.cancerresearchuk.org/cancerstats/types/breast/incidence/index.htm. Accessed 11. January 2012.
142. Seigneurin A, François O, Labarère J, Oudeville P, Monlong J, Colonna M. Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ* 2011;343:d7017.
143. Gøtzsche PC, Jørgensen KJ. Re: Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ* 1. Dec 2011; http://www.bmj.com/rapid-response/2011/12/01/re-overdiagnosis-non-progressive-cancer-detected-screening-mammography-sto. Accessed 11. January 2011.
144. Bennett RL, Sellars SJ, Moss SM. Interval cancers in the NHS breast cancer screening programme in England, Wales, and Northern Ireland. *Br J Cancer* 2011;104:571-7.
145. Perry N, Broeders M, de Wolf C, Törnberg S (Eds.). *European Guidelines for quality assurance in breast cancer screening and diagnosis.* 4th edn. Luxemburg: Offfice for the official publications of the European Communities, 2006.
146. Jensen A, Olsen AH, von euler-Chelpin M, Njor SH, Vejborg I, Lynge E. Do Nonattenders in Mammography Screening Programmes Seek Mammography Elsewhere? *Int J Cancer* 2005;113:464-70.
147. Carlsen K, Høybye MT,Dalton SO, Tjønneland A. Social inequality and incidence of and survival from breast cancer in a population-based study in Denmark. *Eur J Cancer* 2008;44:1996-2002.
148. Christenfeld NS, Sloan RP, Caroll D, Greenland S. Risk factors, confounding, and the illusion of statistical control. *Psychosom Med* 2004;66:868-75.
149. Keen JD, Keen JE. What's the point: will screening mammography save my life? *BMC Med Informatics Decision Making* 2009;9:18.
150. Welch GH, Frankel BA. Likelihood that a womanwith screen-detected breast cancer has had her "life saved" by that screen. *Arch Int Med* 2011;doi:10.1001/archintmed.2011.476.
151. Bock K, Cawson J, Damtjernhaug B et al. Effect of population based screening on breast cancer mortality. *Lancet* 2011; 378: 1775-6.
152. Gøtzsche PC, Jørgensen KJ. Effect of population-based screening on breast cancer mortality. *Lancet* 2012; DOI:10.1016/S0140-6736(12)60039-7.
153. Gärtner R, Jensen M, Nielsen J et al. Prevalence of and factors associated with persistent pain following breast cancer surgery. *JAMA* 2009;302:1985–92.
154. Darby S, McGale P, Peto R et al. Mortality from cardiovascular disease more than 10 years

after radiotherapy for breast cancer: nationwide cohort study of 90 000 Swedish women. *BMJ* 2003;326:256–7.

155. Darby SC, McGale P, Taylor CW,et al. Long-term mortality from heart disease and lung cancer after radiotherapy for early breast cancer: prospective study of about 300 000 women in US SEER cancer registries. *Lancet Oncol* 2005;6:557–65.
156. McGale P, Darby SC, Hall P, et al. Incidence of heart disease in 35,000 women treated with radiotherapy for breast cancer in Denmark and Sweden. *Radiother Oncol* 2011;100:167-75.
157. General Medical Council. *Consent: patients and doctors making decisions together*. London: General Medical Council, 2008. Available at: http://www.gmc-uk.org/guidance/ethical_guidance/consent_guidance/index.asp. Accessed 21 May 2010.
158. [The Danish Health Legislation]. Available at: http://www.im.dk/artikler_im_dk/Files/Fil1/3857.pdf. Accessed 21 May 2010.
159. [The Norwegian Health Legislation]. Available at: http://www.lovdata.no/all/nl-19990702-063.html. Accessed 21 May 2010).
160. Østerlie W, Solbjør M, Skolbekken JA, Hofvind S, Sætnan AR, Forsmo S. Challenges of informed choice in organised screening. *J Med Ethics* 2008;34:e5.
161. Bonfill X, Marzo M, Pladevall M, Martí J, Emparanza JI. Strategies for increasing the participation of women in community breast cancer screening. *Cochrane Database of Systematic Reviews* 2001, Issue 1. Art. No.:CD002943.
162. Coulter A (ed.). Informed choice in screening. *Health Expectations* 2001;4:79-139.
163. Gummersbach E, Piccoliori G, Oriol Zerbe C, Altiner A, Othman C, Rose C, Abholz HH. Are women getting relevant information about mammography screening for an informed consent: a critical apprialsal of information brochures used for screening invitation in Germany, Italy, Spain and France. *Eur J Public Health* 2009;doi:10.1093/eurpub/ckp174.
164. Saalasti-Koskinen U, Mäkelä M, Saarenmaa I, Autti-Rämö I. Personal Invitations for Population-based Breast Cancer Screeening. *Acad Radiol* 2009;16:546-50.
165. Schmidt H, Gerber A, Stock S. What can we learn from German health incentive schemes? *BMJ* 2009;339:b3504.
166. Tuffs A. Germany will penalize cancer patients who do not undergo regular screening. *BMJ* 2006;333:887.
167. Baum M, McCartney M, Thornton H, Bewley S, Pharoah P, Kaplan RM et al. Breast screening peril. Negative consequences of the breast screening programme. *The Times* 19 February 2009. Available at: http://www.timesonline.co.uk/tol/comment/letters/article5761650.ece.
168. Smyth C. NHS rips up breast cancer leaflet and starts all over again. *The Times* 21 February 2009. Available at: http://www.timesonline.co.uk/tol/life_and_style/health/article5776804.ece.
169. Mathieu E, Barratt A, Davey HM, McGeechan K, Howard K, Houssami N. Informed choice in mammography screening: a randomised trial for a decision aid for 70-year-old women. *Arch Int Med* 2007;167:2039-46.
170. Jørgensen KJ. Editorial: Mammography screening is not as good as we hoped. *Maturitas* 2010;65:1-2.
171. Marteau T, Mann E, Prevost AT, Vasconcelos JC, Kellar I, Sanderson S, Parker M, Griffin S, Sutton S, Kinmonth Al. Impact of an informed choice invitation on uptake of screening for diabetes in primary care (DICISION): randomised trial. *BMJ* 2010;340:c2138.
172. Roberts MM. Breast Screening: time for a rethink? *BMJ* 1989;299:1153-55.
173. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2009;151:727-37.
174. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med.* 2011;155:10-20.

175. Keen JD. Analysis of health benefits and cost-effectiveness of breast cancer screening. *Ann Intern Med* 2011; 155:566.
176. Schousboe JT, Kerlikowske K, Cummings SR. In response: Analysis of health benefits and cost-effectiveness of breast cancer screening. *Ann Intern Med* 2011; 155:566-7.
177. Raftery J, Chorozoglou M. Possible net harms of breast cancer screening: updated modelling of Forrest report. *BMJ* 2011; 343:d7627
178. Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman, DG, Schulz K, Moher D: Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-788.
179. Als-Nielsen B, Chen W, Gluud C, Kjærgard LL: Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA* 2003;290:921-928.
180. Jørgensen KJ, Johansen HK, Gøtzsche PC. Flaws in design, analysis an interpretation of Pfizer's antifungal trials of voriconazole and uncritical subsequent quotations. *Trials* 2006;7:3.
181. Jørgensen AW, Jørgensen KJ, Gøtzsche PC. Unbalanced reporting of benefits and harms in abstracts on rofecoxib. *Eur J Clin Pharmacol* 2010;66:341-7.
182. Jørgensen AW, Hilden J, Gøtzsche PC**.** Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. *BMJ* 2006;333:782-5.
183. Barnes DE, Bero LA: Why review articles on the health effects of passive smoking reach different conclusions. *JAMA* 1998;279:1566-70.
184. Peppercorn J, Blood E, Winer E, Partirdge A. Association between pharmaceutical involvement and outcomes in breast cancer clinical trials. *Cancer* 2007;109:1239-46.
185. American Cancer Society. Available at: http://www.cancer.org/docroot/AA/AA_4_Corporate_Recognition.asp. Accessed 25 May 2010.
186. Medical Screening Society. Available at: http://www.medicalscreeningsociety.com/members.asp. Accessed 25 May 2010.
187. Jørgensen KJ, Gøtzsche PC. Mammography screening hasn't lived up to expectations. *BMJ* 2010 Available at: http://www.bmj.com/cgi/eletters/340/mar23_1/c1241#234237. Accessed 25 May 2010.
188. Gøtzsche PC, Jørgensen KJ, Zahl PH. Breast screening: why estimates differ by a factor of 20-25. *J Med Screen* 2010;17:158-9.
189. Jørgensen KJ, Gøtzsche PC. Hand evaluation to NICE. *BMJ* 2011;343:d7533.
190. McPherson K. We need scientific consensus founded on all the evidence. *BMJ* 2011;343:d7529.
191. Baum M. An appeal to Mike Richards. *BMJ* 2011;343:d7535.
192. Kendrick ME. A truly independent review is needed. *BMJ* 2011;343:d7538.
193. Baines CJ. Lessons from US reviews shows impartiality is needed. *BMJ* 2011;343:d7539.
194. Fallowfield LJ. What should women do in the meantime? *BMJ* 2011;343:d7544.
195. Cancer Research UK. See: http://info.cancerresearchuk.org/publicpolicy/ourpolicypositions/symptom_Awareness/cancer_screening/breast-screening-review/. Accessed 6. January 2012.
196. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, Conant F, Fajardo LL, Bassett L, D'Orsi C, Jong R, Rebner M. Digitial Mammograpchic Imaging Trial (DMIST) Investigators Group. *N Engl Jour Med* 2005;353:1773-83.
197. Esserman LJ, Shieh Y, Rutgers EJT et al. Impact of mammographic screening on the detection of good and poor prognosis breast cancers. *Breast Cancer Res Treat* 2011;130:725-34.
198. Baum M. Should routine screening by mammography be replaced by a more selective service of risk assessment/risk management? *Women's Health* 2010;6:71-6.
199. Berrington de Gonzalez A, Berg CD, Visvanathan K, Robson M. Estimated risk of radiation-induced breast cancer from mammographic screening for young BRCA

mutation carriers. *J Natl Camcer Inst* 2010;101:205-9.
200. Borre M, Andersen JR, Iversen P. [*Screening for prostate cancer – what does the evidence and the Danish Urological Society say?*]. Available at: http://www.urologi.dk/rapporter/prostatacancer/PC_screening_sst_2008.pdf. Last accessed 25 May 2010.
201. Ilic D, O'Connor D, Green S, Wilt T. Screening for prostate cancer. *Cochrane Database of Systematic Reviews* 2006, Issue 3. Art. No.: CD004720.
202. Maurizio Infante and colleagues. A randomized study of lung cancer screening with spiral computed tomography. *Am J Respir Crit Care Med* 2009;180:445-53
203. Hewitson P, Glasziou PP, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database of Systematic Reviews* 2007, Issue 1. Art. No.: CD001216.
204. Lentz S. BreastCare 2010. Available at: http://www.breastcare.dk/ombreastcare.html. Last accessed 25 May 2010.
205. Goodman SN. The mammography dilemma: a crisis for evidence based medicine? *Ann Intern Med* 2002;137:363-5.
206. Ioannidis JP, Trikalinos T. Early extreme contradictory estimates may appear in published research: The Proteus phenomenon in molecular genetics research and randomized trials. *J Clin Epidemiol* 2005;58:543-9.
207. Chalmers I, Matthews R. What are the implications of optimism bias in clinical research? *Lancet* 2006;367:449-50.

DANISH SUMMARY

Rationalet for brystkræftscreening med mammografi er besnærende enkelt: Find brystkræften tidligt og reducér dødeligheden og behovet for brystfjernelser. Men brystkræft er en kompleks sygdom, og komplicerede problemer har sjældent enkle løsninger.

Brystkræftscreening fremrykker diagnosetidspunktet beskedent i forhold til knudens samlede levetid, og screening-detekterede tumorer har stadig en størrelse, hvor spredning kan forekomme. Et centralt spørgsmål er, om screening kan forebygge spredningen og mindske behovet for totale brystfjernelser, til fordel for brystbevarende operationer.

En reduktion af dødelighed kan aldrig i sig selv retfærdiggøre en medicinsk intervention, men skal holdes op imod skadevirkningerne. For mammografiscreening er overdiagnostik den vigtigste skadevirkning, og i de seneste år er der kommet en bredere erkendelse af dette. Screening fører til diagnosticering og behandling af brystkræft, som ellers aldrig ville være opdaget i kvindens levetid, fordi den vokser langsomt, eller slet ikke. Screening gør derfor kvinder til kræftpatienter uden grund, med livslange negative psykiske og fysiske konsekvenser. Debatten er således langt mere nuanceret end et simpelt spørgsmål om, hvorvidt screening reducerer dødeligheden af brystkræft.

I denne afhandling opgøres de vigtigste gavnlige og skadelige virkninger af mammografiscreening. Danmark har en uscreenet 'kontrolgruppe', fordi kun to regioner tilbød screening gennem en lang periode, hvilket er unikt i international sammenhæng. Dette blev udnyttet til at opgøre effekten på brystkræftdødelighed, omfanget af overdiagnostik og brug af totale brystfjernelser. Sammen med et systematisk review der inkluderede fem andre lande, kunne vi vise at omkring halvdelen af screeningsdetekteret brystkræft er overdignosticeret. En effekt på brystkræftdødeligheden er tvivlsom i dag, og overdiagnostik medfører en stigning i anvendelsen af totale brystfjernelser. Disse fund diskuteres i forhold til kræftbiologi og stadieinddeling.

Den information som kvinder får, når de inviteres til mammografiscreening, og som findes på Internettet, overdriver de gavnlige virkninger, og der opfordres direkte til deltagelse. Skadevirkningerne nedtones eller udelades, selvom der er enighed om, at målet er et informeret samtykke. Dette rejser en etisk diskussion om autonomi og paternalisme, og om vanskelighederne ved at afveje fordele mod ulemper.

Endelig diskuteres betydningen af økonomiske, politiske og professionelle interessekonflikter, samt sundhedsøkonomiske forhold.

PAPERS INCLUDED IN THIS THESIS

Listed as they first appear in the text.

1. Jørgensen KJ, Gøtzsche PC. Presentation on web sites of possible benefits and harms from screening for breast cancer: cross sectional study. *BMJ* 2004; 328: 148-51.
2. Jørgensen KJ, Gøtzsche PC. Content of invitations for publicly funded screening mammography. *BMJ* 2006; 332: 538-541.
3. Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark. A comparative study. *BMJ* 2010; 340: c1241.
4. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009; 339: b2587.
5. Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009; 9: 36.

Information in practice

Presentation on websites of possible benefits and harms from screening for breast cancer: cross sectional study

Karsten Juhl Jørgensen, Peter C Gøtzsche

Abstract

Objective To investigate whether information on mammographic screening presented on websites by interest groups is balanced, is independent of source of funding, and reflects recent findings.

Design Cross sectional study using a checklist with 17 information items.

Setting 27 websites in Scandinavian and English speaking countries.

Results The 13 sites from advocacy groups and the 11 from governmental institutions all recommended mammographic screening, whereas the three from consumer organisations questioned screening (P = 0.0007). All the advocacy groups accepted industry funding, apparently without restrictions. In contrast the three consumer organisations acknowledged the risk of bias related to industry funding, and two of them did not accept such funding at all. Advocacy groups and governmental organisations favoured information items that shed positive light on screening. The major harms of screening, overdiagnosis and overtreatment, were mentioned by only four of these groups, but by all three sites from consumer organisations (P = 0.02). In addition, the chosen information was often misleading or erroneous. The selection of information items for websites did not reflect recent findings, apart from the consumer sites, which were much more balanced and comprehensive than other sites (median of 9 information items *v* 3 items, P = 0.03).

Conclusions The information material provided by professional advocacy groups and governmental organisations is information poor and severely biased in favour of screening. Few websites live up to accepted standards for informed consent such as those stated in the General Medical Council's guidelines.

Introduction

Women can get information about the possible benefits and harms of mammographic screening from governmental institutions and professional advocacy groups. This information could be biased, however, since the success of a screening programme depends on the participation rate. Another potential conflict of interest is industry funding of advocacy groups.

A review of 58 Australian pamphlets in 1998 showed that the information presented to women invited for breast cancer screening was biased and insufficient and did not allow fully informed consent.[1] Another Australian study, of 54 publications used to inform about screening mammography in New South Wales, showed that only 18% of the publications gave any information on false positive and false negative results, and only 48% gave any information on adverse effects.[2]

In the European Union an average of 23% of the population use the internet to find information about health issues; Denmark has the highest rate, at 47%.[3] If the information about screening on the internet is biased, women's status as autonomous individuals could be violated.[4] The importance of balanced information is underlined by a study which found that 61% of women decided for themselves whether to have a screening mammogram, and a further 26% made the decision together with their doctor.[5]

In 2001 the quality of the randomised trials of mammographic screening was criticised in a comprehensive Cochrane review that questioned the benefit of screening.[6] In addition, important harms related to overdiagnosis and overtreatment were demonstrated.[7 8] We therefore decided to study whether the current information on the internet was balanced and reflected the recent findings.

Materials and methods

We studied whether the information presented on the internet by major interest groups gave a balanced account of the possible benefits and harms of mammographic screening; whether funding of interest groups was related to type of information; and whether the information was different from what was previously provided in pamphlets.[1]

We located websites produced by professional advocacy groups (such as cancer charities), governmental institutions, and consumer organisations from Australia, Canada, Denmark, New Zealand, Norway, Sweden, the United Kingdom, and the United States. All of these countries have screening programmes, although so far only regionally in Denmark and Norway. We searched for "breast cancer" *and* "mammography" *and* "screening" *and* one of the included countries, primarily with the search engines Google and Yahoo (see appendix on bmj.com for details). By using these major search engines, we hoped to find the same websites that women would on the internet. The most popular sites with the best match to the search terms top the lists. We located the major organisations in each country with these search engines, and we identified extra websites by links and personal contacts.

One of the authors (KJJ) searched the sites systematically and printed all relevant information. When sites had their own search machine or site map, these were used to locate relevant pages. The printout of each website was evaluated independently

P+ *Details of the search engines used and the websites reviewed in this study appear on bmj.com*

by each author, and any disagreements were settled by discussion. We used a data sheet that contained the same 10 information items as in the review of pamphlets[1] and seven additional relevant items recommended by the Ethical Council in Denmark.[9] When a site had separate information for professional healthcare workers and the general public, we evaluated the information for the public. Scientific articles linked to a site were not evaluated. When information on funding was unclear, we contacted the organisations. We accessed all the sites during September and October 2002.

We divided the websites into three groups, primarily based on the information provided in the "About us" section, according to whether they were from governmental institutions, professional advocacy groups, or consumer organisations. We defined advocacy groups as those whose general purpose is to promote the interests of patients and their relatives and consumer organisations as those whose general aim is to assess the quality of the healthcare services that are offered to patients and citizens.

We hypothesised that the information provided by consumer organisations would be more comprehensive and would more often give information on possible harms than the two other types of organisations. We compared the number of information items provided with a Mann-Whitney test, and individual items with Fisher's exact test.

Results

Recommendations on websites

We located 27 websites, 13 from professional advocacy groups, 11 from governmental institutions, and three from consumer organisations (see appendix on bmj.com). The governmental and advocacy sites all recommended mammography screening, at least implicitly, whereas the consumer sites questioned the value of screening (P = 0.0007).

Funding

All 13 advocacy groups accepted sponsorship from industry, apparently without restrictions. The Canadian Cancer Society noted that "Partnership with the Canadian Cancer Society can assist your company in reaching your commercial objectives." In contrast, the three consumer organisations explicitly acknowledged the risk of bias related to industry funding: two (Breast Cancer Action and Center for Medical Consumers) said that they did not accept grants from industry, while the third (National Breast Cancer Coalition) noted that only 15% of its budget can come from corporations, only 5% from any single source, and that this funding is restricted to general operating support.

Information items

The sites had a median of three information items out of the 17 possible; the highest number was 13 (Center for Medical Consumers). Five sites had none, and these sites mainly addressed practical issues related to the examination. The median number of items was nine for the three consumer sites, which were sceptical about screening, and three for the other sites (P = 0.03). Our two independent assessments of the sites identified a total of 98 and 99 information items; after discussion, we agreed on 118 items. The discrepancies were mainly caused by oversight. The significant difference between the consumer sites and the other sites persisted for the individual assessments (P = 0.03).

The four most common information items were the same as in the 1998 study of pamphlets (table 1), but more websites described the relative and absolute risk reduction of death from breast cancer (P = 0.006 and P = 0.005, respectively), the proportion of women recalled (P = 0.006), and the predictive value of a positive mammogram (P = 0.02). The relative risk reduction was usually given as 30%, but estimates varied from none to 50% reduction. Three times as many sites provided the

Table 1 Presence of information items about screening for breast cancer on 27 websites (from professional advocacy groups, governmental institutions, and consumer organisations) and in a 1998 survey of 58 pamphlets[1]

	No of sites mentioning information item				Occurrence (%)	
Information items	Advocacy sites (n=13)	Governmental sites (n=11)	Consumer sites (n=3)	Total (n=27)	On websites	In 58 pamphlets[1]
Included in 1998 review of 58 pamphlets[1]						
Lifetime risk of developing breast cancer	5	6	1	12	44	60
Lifetime risk of dying from breast cancer	1	2	1	4	15	2
Survival from breast cancer	1	1	1	3	11	5
Relative risk reduction of death from breast cancer	5	7	3	15	56	22
Absolute risk reduction of death from breast cancer	1	2	2	5	19	0
Number needed to screen to avoid one death from breast cancer	0	0	2	2	7	0
Proportion of screened women who would be recalled	6	4	2	12	44	14
Proportion of breast cancers detected by mammography (sensitivity)	2	3	2	7	26	26
Proportion of women without breast cancer who would have a negative mammogram (specificity)	0	0	0	0	0	0
Proportion of women with a positive mammogram who would have breast cancer (positive predictive value)	2	1	1	4	15	0
Added in this study						
Relative risk reduction of total mortality	0	1	1	2	7	
Carcinoma in situ	4	3	3	10	37	
Overdiagnosis and overtreatment	2	2	3	7	26	
Effect of screening on number of mastectomies or lumpectomies	1	4	2	7	26	
Risks related to radiotherapy	1	2	1	4	15	
Psychological distress related to false positive results	4	3	3	10	37	
Pain at mammography	8	5	1	14	52	

relative risk reduction as provided the absolute risk reduction (table 1).

For the seven new items we added to those used in the survey of pamphlets, information was rarely provided on relative risk reduction of total mortality (only two sites did so) and risks related to radiotherapy (four sites). Information on the other items was provided by a quarter to half of the websites (table 1). The three consumer sites mentioned overdiagnosis and overtreatment, but only four of the other 24 sites did so ($P = 0.02$).

Bias in selection and presentation of information

The essence of the messages varied widely (see box). Most websites omitted information on important harms (table 1) and emphasised possible benefits in a way that would be expected to increase uptake of screening. For example, 12 sites mentioned lifetime risk of developing breast cancer, usually followed by the annual number of diagnoses. In contrast, only three sites mentioned the relatively reassuring message that women have a more than 50% chance of surviving breast cancer once it is diagnosed, and only four stated that the lifetime risk of dying from breast cancer is about 3-4% (depending on country). Twelve sites stated the number of women recalled and presented this as about 5% at each screening round.

Issues related to carcinoma in situ, overdiagnosis and overtreatment, and number and type of operations were mentioned by a quarter to a third of the sites (table 1), but often in a misleading or erroneous fashion (see box). Four governmental websites and one advocacy site indicated that screening leads to fewer mastectomies. One governmental and three advocacy sites noted that it is beneficial to detect and remove carcinoma in situ since it would then not recur. Only two such sites mentioned that screening can detect cancers that may never progress, compared with all three consumer sites ($P = 0.007$). Only four sites noted that there could be risks associated with radiotherapy, but the risks were downgraded on three of the sites (see box).

The three consumer sites described psychological distress related to false positive findings, compared with seven of the governmental or advocacy sites ($P = 0.08$); seven sites described it vaguely as "anxiety," and no sites gave an estimate of the incidence. The potential pain inflicted by the mammographic procedure was mentioned by 14 sites, three of which claimed that the procedure shouldn't be painful.

Discussion

The material about screening for breast cancer that was provided by professional advocacy groups and governmental organisations was information poor and severely biased in favour of screening. The material provided by the consumer organisations was much more comprehensive and balanced. It is worrying that so few websites live up to accepted standards for informed consent[12] since it is possible to persuade people to accept or decline cancer screening by withholding or including particular information items.[13 14]

The way data are presented can also substantially affect views on therapeutic effectiveness.[15] To mention that screening reduces the risk of dying from breast cancer by 30%[16] (relative risk reduction) is much more impressive than the equivalent finding—reported in the same overview—that the absolute risk of dying from breast cancer is reduced by 0.1% after 10 years.[16] On the Danish Cancer Society site, this estimate was increased by a factor of 10 to 1%. The inflated estimate also appears in a Danish governmental report that recommended screening be introduced, and in which the results from the screening trials had been extrapolated far beyond the actual data until the women became 80 years of age, assuming they did not die from other causes and disregarding the fact that only women aged 50-69 are invited to screening.[17]

Breast cancer mortality is a biased outcome that favours mammographic screening,[6–8 18] and screening increases cardiovascular mortality because of the increased use of radiotherapy.[8 11] Total mortality is therefore relevant, but only the National Breast Cancer Coalition noted that an effect on total mortality has not been demonstrated.[6–8 19 20] Since misclassification of cause of death often occurs with deaths from other cancers,[6–8] it is also relevant that the screening trials failed to find an effect on death from any cancer (including breast cancer), contrary to what is expected from the claimed 30% reduction in breast cancer mortality.[21] This was not mentioned by any site.

Overdiagnosis and overtreatment

The most important harms of screening—overdiagnosis and overtreatment—seem to be the best kept secret about screening. The level of overdiagnosis can be studied reliably in those two screening trials that were not flawed and were not contaminated by early, systematic screening of the control group.[22 23] It was 30% (table 2), which corresponds to the 31% overtreatment previously reported.[7 8] The overdiagnosis was 33% for the Swedish trials that screened the whole control group when only cancers before this screen were included.[24]

Epidemiological data show similar levels of overdiagnosis. When screening was introduced in the United States, the incidence of invasive breast cancer increased by 26% in only seven years and has remained elevated ever since over a span of 20 years.[24 25] If carcinoma in situ cases are added, the overdiagnosis increases to about 35%.[24 25] In the United Kingdom the incidence of invasive cancer increased by 37% and cases of carcinoma in situ increased by 373% from 1990 to 2001, when screening was introduced.[26]

These results indicate that the five websites that noted that screening leads to fewer mastectomies are seriously misleading. The opposite seems to occur. In the screening trials 20% more mastectomies were performed in the screened groups than in the control groups,[7 8] and in the United Kingdom mastectomies increased by 36% for invasive cancer and by 422% for carcinoma in situ when screening was introduced.[25] Because of overdiagnosis, screening also increases the use of radiotherapy,[8] but only four sites gave the important information that radiotherapy was associated with risks,[11] and three of the four sites downgraded this information (see box).

Downgrading of other harms

The websites' statements that about 5% of screened women would be recalled at each screening round is far less disturbing than the information that the cumulated risk is 49% after 10 mammograms.[10] The information that false positive findings can sometimes create "anxiety" is also much more soothing than the information that more than 10% of women screened will at some point experience important psychological distress for many months.[6 8]

The websites generally downgraded the potential pain inflicted by the mammographic procedure, and the claim by three websites that "the procedure shouldn't be painful" is highly misleading. A survey of five studies found that 31% of women felt pain during their first mammogram and that a further 23% felt it was very uncomfortable.[27] Furthermore, half of 81 women who declined an invitation to the second round of

Comments on possible harms from breast cancer screening by websites

Carcinoma in situ

"There is no advantage to early detection of [carcinoma in situ]; in fact, there is a huge disadvantage of unnecessary treatment"—Breast Cancer Action

"Women in this situation sometimes have more extensive surgery than women with invasive cancer"—BreastScreen Aotearoa

"There is virtually no risk of the cancer coming back once it has been removed"—Cancer Research UK

"These early tumors cannot harm patients if they are removed at this stage and mammography is the only proven method to reliably detect these tumors"—RadiologyInfo

Overdiagnosis and overtreatment

"Regular mammography screening may actually increase a woman's chances of losing a breast … Mammograms find some early cancers that might never have been diagnosed and some of these early cancers are treated by mastectomy"—Center for Medical Consumers

"We cannot determine at the time of diagnosis the type of tumor a woman has. The result is that we mistreat or over-treat many women diagnosed with breast cancer in our effort to help the others"—Breast Cancer Action

"There are some types of early breast cancers that will never spread to other parts of the body, and mammograms probably find many of these breast cancers … The result is that many women get treated for breast cancer when they may not need to be treated at all … This is called 'overtreatment'. Overtreatment can be harmful to women"—National Breast Cancer Coalition

"Over-diagnosis and over-treatment are estimated to account for between 0-10% of cancers detected by breast screening"—BreastScreen Aotearoa (Our comment: it amounts to 30% (see table 2 and text))

"Screening detects primarily those early changes which will later develop into cancer [our translation]"—Kræftens Bekæmpelse (Our comment: this is not true for carcinoma in situ)

"Only in 11% of false positive cases was it necessary to remove the lump in order to exclude the suspicion of cancer [our translation]"—Kræftens Bekæmpelse (Our comment: since the proportion false positives is about 5% at each screen, the quote gives the impression that only 0.5% of screened women get an unnecessary biopsy. However, over 10 mammograms, the cumulated biopsy risk for those screened was estimated to be 19%,[10] or 40 times higher)

Numbers of mastectomies and lumpectomies

"Mammography screening leads to more false-positives, more unnecessary surgeries, and more use of aggressive breast cancer treatments … Mammography screening also increased the number of mastectomies by 20% and the number of mastectomies and lumpectomies combined by 30%"—National Breast Cancer Coalition

"Treating breast cancer when it is small … increases the likelihood that she can be offered surgical options which conserve the breast"—BreastScreen Aotearoa

"With early detection the need for radical surgery or radiation therapy, with their adverse side effects, can be minimized"—Health Canada-Womens Health Bureau

Risks related to radiotherapy

"Women may undergo unnecessary and/or inappropriate treatments … chemotherapy and radiotherapy are toxic and should not be given to women who do not need them"—National Breast Cancer Coalition

"Because the current technique of radiotherapy defines both doses and target volume precisely, the doses to healthy near-by tissues are minimal. A Danish study with 12 years of follow-up did not find an increase in heart disease after radiotherapy (Hojris et al, *Lancet* 1999). The claims that current radiotherapy of breast-cancer patients causes heart disease are therefore not correct [our translation]"—Kreftregistret (Our comment: this study had too little power and too short a follow up to exclude this possibility (11 *v* 11 vascular deaths), and a systematic review of radiotherapy indicated that in low risk women, such as those with cancers found by screening, it would be expected to increase mortality from all causes[11])

"There are other more severe, but much rarer long-term side effects. Many of these don't happen now. This is because the treatment planning is much more exact"—Cancer Research UK (Our comment: there was no indication that the subject discussed was radiotherapy and no mention of the finding that radiotherapy can cause fatal cardiovascular disease)

Pain at mammography

"Most women find this uncomfortable and some find it painful"—BreastScreen Aotearoa

"May cause some temporary pain, but it is usually not severe"—Health Canada-Womens Health Bureau

"They are so satisfied that they recommend fellow sisters to get x-rayed in this way. They do this although two out of three experience the examination as slightly uncomfortable (52.3%), hurting (11.4%) or painful (2.5%) [our translation]"—Den Norske Kreftforening

"You should not feel pain"—American Cancer Society

"Fast and almost painless examination [our translation]"—Cancerfonden

"Shouldn't be painful"—Cancer Research UK

Other citations

"Show up when invited for population based control (screening) [our translation]"—Den Norske Kreftforening

"The leaflet includes an explanation about false positive and false negative results"—NHS Cancer Screening Programmes (Our comment: there is no such explanation in the leaflet)

"If the programme is to be successful for women it is very important that you return for screening every two years"—BreastScreen Aotearoa (Our comment: the focus is on the programme, not on the women)

"Women in the 50-69 target group who have regular two-yearly mammograms can reduce their risk of dying from breast cancer by up to 50%"—BreastScreen Australia (Our comment: this percentage is seriously misleading[6–8])

"Early detection of breast cancer saves lives. Therefore, there is no need to re-evaluate the value of general mammography screening [our translation]"—Cancerfonden

Table 2 Numbers of cancers detected, and lumpectomies and mastectomies in the trials of breast cancer screening from Canada (after 7 years)[22] and Malmö, Sweden (after 8.8 years)[23]

Trial	No of women		No of cancers detected			No of lumpectomies and mastectomies		
	Screened group	Control group	Screened group	Control group	Relative risk (95% CI)	Screened group	Control group	Relative risk (95% CI)
Canada[22]:								
Women aged 40-49	25 214	25 216	426	327	1.30 (1.13 to 1.50)	415	313	1.33 (1.15 to 1.53)
Women aged 50-59	19 711	19 694	460	365	1.26 (1.10 to 1.44)	448	351	1.28 (1.11 to 1.46)
Malmö[23]	21 088	21 195	588	447	1.32 (1.17 to 1.49)	561	419	1.34 (1.18 to 1.52)
Overall					1.30 (1.20 to 1.40)			1.31 (1.22 to 1.42)

CI=confidence interval.

screening said that their major reason for doing so was because their first mammogram was painful.[28]

Potential conflicts of interest

The three consumer organisations, which all questioned the value of screening, did not have an apparent conflict of interest. In contrast, all advocacy groups accepted financial support from industry, apparently without restrictions. Receiving financial support from companies with an economic interest in screening programmes, or the treatments associated with them, potentially undermines the objectivity of the organisation, with a mutually beneficial relationship as a possible outcome.[29 30]

Governmental organisations that offer screening are also potentially biased. They have made a decision and must defend this position.[4] Furthermore, the success of screening depends on a high participation rate to make the programmes fit the preceding cost-benefit analyses. This bias can result in complete misrepresentation of unwelcome research results.[31]

Had the websites selected the most important information items?

Six of the seven information items we added to the previous survey of pamphlets addressed possible harms of screening (table 1), although some of them were sometimes described as benefits. Sites with a positive presentation of screening could therefore seem relatively more information poor. However, this cannot explain why most sites downgraded the harms that they did mention and omitted information that is vital to make an informed decision.[32] Our study shows that it is not the relative importance of the information items that determines whether they are mentioned. The risk of becoming a cancer patient unnecessarily and the increased risk of loosing a breast[8 26] are obviously important for informed decision making but were rarely mentioned.

It could be argued that some women would have trouble understanding the meaning of an absolute risk reduction. But it is not more difficult to understand than a relative risk reduction, and it is more important for informed decision making,[15] in particular when the event rate is low, as it is for breast cancer mortality. Nor is it difficult to understand that screening can detect harmless tumours.

Tension between informed consent and high uptake

In accordance with policies of national screening programmes,[33] most sites stated that women's decision whether to participate should be based on informed consent. Requirements for informed consent should be stricter when the healthy population is approached than when a sick patient consults a doctor, since healthy people have not asked for help and are considering participation in tests on a different basis. For breast screening, however, our healthcare systems have done the opposite and have sacrificed the obligation of a fully informed consent for a paternalistic role, as shown in the NHS leaflet that asks "Why do I need breast screening?" rather than "Do I need breast screening?"[34] If the concern is, as screening advocates have suggested, that too few women would participate if they were presented with the relevant issues,[4 35] screening may be too controversial to be justifiable.

The bottom line of mammography screening

The effect of screening is uncertain since most trials are of poor quality.[6–8] The most optimistic and most quoted result is a 30% reduction in breast cancer mortality.[16] If it were true, it would mean that one woman would be saved from dying from breast cancer for every 1000 women invited to screening for 10 years.[16] After 10 years of screening, 90.3% of the women would be alive, whereas if they were not screened 90.2% would be alive.[16 36] However, it is also possible that no one will be saved, since the women may die from something else, such as from complications from the breast cancer treatment,[8 11] and since an effect of screening on mortality from all causes has not been demonstrated.[6–8 19 20]

The overdiagnosis means that for every 1000 women invited to screening for 10 years, five additional women will be diagnosed with cancer; two additional women will have a breast removed and three will have a lump removed.[8]

Thus, most optimistically, for every woman who has her life prolonged, five healthy women, who would not have received a breast cancer diagnosis if there had not been screening, will be converted into cancer patients. Whether this is a too high price to pay is open to debate, but if women and policy makers are not informed of this balance between major benefits and major harms—which they have not been so far—then there cannot be any discussion or rational decision making. The present situation is that a woman customer who visits a "screening shop" doesn't know what she is buying into, and most often the shopkeeper either doesn't know or doesn't tell. This is untenable.

Suggested improvements in the information women are offered

It is inappropriate to continue to use information about screening purely for encouraging high uptake.[4] Whatever is presented should be balanced and should reflect fairly the level of scientific uncertainty, allowing women to reach a decision by themselves. Furthermore, there should be links to more detailed information for those who need it.

Possible benefits and harms should get similar attention and should be presented in a similar fashion. If, for example, the benefit is presented as a 30% reduction in breast cancer mortality, then the overdiagnosis and overtreatment should be given as a 30% increase. However, it would be preferable to use absolute risk reductions and numbers[15]—for example, most optimistically, one woman has her life prolonged for every five women who get an unnecessary cancer diagnosis and treatment. The fact that

What is already known on this topic

A 1998 survey showed that the information material in pamphlets presented to Australian women invited for breast cancer screening was biased and insufficient, and did not allow fully informed consent

What this study adds

In 2001 the quality of the randomised trials of mammographic screening was criticised in a comprehensive systematic review, which questioned the benefit of screening and documented important harms

Despite these findings, the information presented to women on websites by professional advocacy groups and governmental organisations was selective and biased and failed to mention major harms

Websites from consumer groups were more balanced and comprehensive than sites by professional advocacy groups and governmental organisations

screening can detect cancers that may never progress is little known to women and should be emphasised in information material.[32] It also seems relevant to note that the detection of carcinoma in situ can cause problems for women applying for medical or life insurance, even for their daughters.[37]

The symmetry of information should also be respected for cumulated risks. If the lifetime risk of getting breast cancer is noted, then the lifetime risk of getting a false positive diagnosis should also be noted rather than the risk at each screening round.

Since people should be informed about the uncertainties of screening,[12] they need to know that the effect of screening is uncertain. A recent review by the US Preventive Services Task Force gave mammography screening a grade B recommendation, which means: "Evidence is sufficient to determine effects on health outcomes, but the strength of the evidence is limited by the number, quality, or consistency of the individual studies; generalizability to routine practice; or indirect nature of the evidence on health outcomes."[38]

We will send a copy of this article to the organisations whose websites we surveyed.

We thank Lacie Koppelman (science analyst, National Breast Cancer Coalition), Mary Ann Napoli (director, Center for Medical Consumers), and Marjukka Mäkelä (director, Finnish branch of the Nordic Cochrane Centre, and director, Finnish Office for Health Care Technology Assessment) for helpful advice and encouragement.The views expressed in this article represent those of the authors and are not necessarily the views or the official policy of the Cochrane Collaboration.

Contributors: PCG conceived the project, KJJ wrote the draft protocol, did the searches, and wrote the first manuscript. Both authors extracted data and contributed to the protocol and the manuscript. Both are guarantors.

Funding: None.

Competing interests: One of the authors was involved in the systematic review of breast screening trials that questioned the value of screening.

1 Slaytor EK, Ward JE. How risks of breast cancer and the benefits of screening are communicated to women: analysis of 58 pamphlets. *BMJ* 1998;317:263-4.
2 Croft E, Barratt A, Butow P. Information about tests for breast cancer: what are we telling people? *J Fam Pract* 2002;51:858-60.
3 European Opinion Research Group. European Union citizens and sources of information about health. 2003. http://europa.eu.int/comm/health/ph_information/documents/eb_58_en.pdf (accessed 3 Oct 2003).
4 Raffle AE. Information about screening—is it to achieve high uptake or to ensure informed choice? *Health Expect* 2001;4:92-6.
5 Davey HM, Barratt AL, Davey E, Butow PN, Redman S, Houssami N, et al. Medical tests: women's reported and preferred decision-making roles and preferences for information on benefits, side-effects and false results. *Health Expect* 2002;5:330-40.
6 Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2001;(4):CD001877.
7 Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001;358:1340-2.
8 Olsen O, Gøtzsche PC. Systematic review of screening for breast cancer with mammography. 2001. http://image.thelancet.com/lancet/extra/fullreport.pdf (accessed 8 Jan 2004).
9 *Screening—en redegørelse*. København: Det Etiske Råd; 1999.
10 Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338:1089-96.
11 Early Breast Cancer Trialists' Collaborative Group. Favourable and unfavourable effects on long-term survival of radiotherapy for early breast cancer: an overview of the randomised trials. *Lancet* 2000;355:1757-70.
12 General Medical Council. *Seeking patients' consent: the ethical considerations*. London: General Medical Council, 1998. www.gmc-uk.org/standards/consent.htm (accessed 8 Jan 2004).
13 Wolf AM, Nasser JF, Wolf AM, Schorling JB. The impact of informed consent on patient interest in prostate-specific antigen screening. *Arch Intern Med* 1996;156:1333-6.
14 Ransohoff DF, Harris R. Lessons from the mammography screening controversy: can we improve the debate? *Ann Intern Med* 1997;127:1029-34.
15 Naylor D. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness. *Ann Intern Med* 1992;117:916-21.
16 Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993;341:973-8.
17 *Tidlig opsporing og behandling af brystkræft: statusrapport*. København: Sundhedsstyrelsen, 1997.
18 Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and surgery in early breast cancer: an overview of the randomized trials. *N Engl J Med* 1995;333:1444-55.
19 Gøtzsche PC. Update on effects of screening mammography. *Lancet* 2002;360:338.
20 Nyström L, Andersson I, Bjurstam N, Frisell J, Rutqvist LE. Update on effects of screening mammography: authors' reply [correction appears in *Lancet* 2002;360:1178]. *Lancet* 2002;360:339-40.
21 Gøtzsche PC. Screening for breast cancer with mammography: author's reply. *Lancet* 2001;358:2167-8.
22 Miller AB. The costs and benefits of breast cancer screening. *Am J Prev Med* 1993;9:175-80.
23 Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ* 1988;297:943-8.
24 Gøtzsche PC. On the benefits and harms of screening for breast cancer. *Int J Epidemiol* (in press).
25 Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, et al, eds. *SEER cancer statistics review, 1973-1999*. Bethesda, MD: National Cancer Institute, 2002. (http://seer.cancer.gov/csr/1973_1999/).
26 Douek M, Baum M. Mass breast screening: is there a hidden cost? *Br J Surg* 2003;90(suppl 1):June (Abstract Breast 14).
27 McNoe B, Miller D, Elwood M. *Women's experience of the Otago-Southland breast screening programme—a compilation of five studies*. Dunedin: Hugh Adam Cancer Epidemiology Unit, 1996.
28 Elwood M, McNoe B, Smith T, Bandaranayake M, Doyle TC. Once is enough—why some women do not continue to participate in a breast cancer screening programme. *N Z Med J* 1998;111:180-3.
29 Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003;289:454-65.
30 Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167-70.
31 Gøtzsche PC. Office of NHS cancer screening programme misrepresents Nordic work in breast screening row. *BMJ* 2001;323:1131.
32 Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *BMJ* 2000;320:1635-40.
33 Mateau TM. A measure of informed choice. *Health Expect* 2001;4:99-108.
34 NHS. *Breast screening, the facts*. London: Health Promotion England, 2001. www.cancerscreening.nhs.uk/breastscreen/publications/ia-02.html (accessed 8 Jan 2004).
35 Horton R. Screening mammography: setting the record straight. *Lancet* 2002;359:441-2.
36 Nyström L, Larsson LG, Wall S, Rutqvist LE, Andersson I, Bjurstam N, et al. An overview of the Swedish randomised mammography trials: total mortality pattern and the representivity of the study cohorts. *J Med Screen* 1996;3:85-7.
37 Thornton H. Consequences of breast screening. *Lancet* 2000;356:1033.
38 US Preventive Services Task Force. Screening for breast cancer: recommendations and rationale. *Ann Intern Med* 2002;137:344-6.

(Accepted 9 November 2003)

bmj.com 2004;328:148

Nordic Cochrane Centre, H:S Rigshospitalet, DK-2100 København Ø, Denmark
Karsten Juhl Jørgensen *physician*
Peter C Gøtzsche *director*

Correspondence to: Peter C Gøtzsche pcg@cochrane.dk

Appendix: search engines used and websites reviewed

Search engines used

- Google.com
- Jubii.dk
- msn.se
- Sol.no
- Yahoo.com

Reviewed websites

Advocacy groups

- American Cancer Society (www.cancer.org). Mammography and other breast screening procedures
- Breast Cancer Care (www.breastcancercare.org.uk). Screening for breast cancer
- Canadian Cancer Society (www.cancer.ca). Early detection and sceening of breast cancer
- Cancer Council Australia (www.cancer.org.au)
- Cancerfonden (www.cancerfonden.se). Mammografi
- Cancer Research UK (www.cancerhelp.org.uk)
- Den Norske Kreftforening (www.kreft.no)
- Kræftens Bekæmpelse (www.cancer.dk). Mammografi—undersøgelse for brystkræft
- New Zealand Breast Cancer Foundation (www.nzbcf.org.nz)
- RadiologyInfo (www.radiologyinfo.org). Mammography
- Susan G. Komen Breast Cancer Foundation (www.komen.org). The ABC's of breast cancer
- Y-ME (www.y-me.org). Mammography—the way to earlier detection
- Womens Health Network (www.womenshealthnetwork.org). Breast cancer detection and diagnosis

Consumer organisations

- Breast Cancer Action (www.bcaction.org)
- Center for Medical Consumers (www.medicalconsumers.org) and link from this site to The University of California at San Francisco (http://mammography.ucsf.edu) (use "Links", "Cancer").
- National Breast Cancer Coalition (www.stopbreastcancer.org). The mammography screening controversy: Questions and answers

Governmental organisations

- BreastScreen Aotearoa. Ministry of Health, New Zealand (www.healthywomen.org.nz)
- BreastScreen Australia. Australian Department of Health and Ageing (www.breastscreen.info.au)
- Breast Test Wales (www.velindre-tr.wales.nhs.uk/btw/index.html). Questions and answers about breast screening
- Federal Drug Administration (www.fda.gov). Mammography today
- Folkehelseinstituttet (www.folkehelsa.no). Mammografi og mammografiprogrammet

- Health Canada—Womens Health Bureau (www.hc-sc.gc.ca). Mammography and women's health
- Kreftregisteret (www.kreftregisteret.no). Mammografiprogrammet
- National Breast Cancer Centre—Australia (www.nbcc.org.au)
- National Cancer Institute—USA (www.cancer.gov). Breast cancer: screening and screening mammograms: questions and answers
- NHS Direct (www.nhsdirect.nhs.uk). NHS cancer screening programmes. Breast screening—the facts
- Socialstyrelsen (www.socialstyrelsen.se). Mammografi—frågor och svar

Abstract of this article
Full Text of this article
Email this article to a friend
Respond to this article
Alert me when:
New articles cite this article

Home Help Search/Archive Feedback Table of Contents

Analysis and comment

Public health

Content of invitations for publicly funded screening mammography

Karsten Juhl Jørgensen, Peter C Gøtzsche

The benefits and harms of screening for breast cancer are delicately balanced and women should decide for themselves, on an informed basis. Do the invitations give enough information to enable this?

See also p 499

Nordic Cochrane Centre, Rigshospitalet Department 7112, Blegdamsvej 9, DK-2100 Copenhagen, Denmark

Karsten Juhl Jørgensen *research fellow*

Peter C Gøtzsche *director*

Correspondence to: K J Jørgensen kj@cochrane.dk

BMJ 2006;332:538–41

Invitations to screening mammography play a central part in the process of obtaining informed consent. It is the only source of information distributed to all potential participants. Other sources, such as pamphlets and websites, have been shown to be information poor and biased in favour of participation,[1 2 w1 w2] and information from the media and doctors is likely to vary and be unevenly distributed. We examined mammography invitations from English speaking and Scandinavian countries with publicly funded screening to assess whether they provide sufficient information to enable women to make an informed decision.

Information versus high uptake

When a society decides to offer cancer screening, eligible citizens need to be made aware of the programme. A letter of invitation is a common approach, and it seems obvious to use this letter to provide balanced information about benefits and harms of screening, particularly since there is international consensus that participation in cancer screening should be based on informed consent.[3 w3 w4] However, in countries with publicly funded screening, those responsible for the success of the programme are also those who provide the information. Herein lies a potential conflict of interest. High participation rates are pivotal to any screening programme, but information about potential harms may deter women from participation.

GARO/PHANIE/REX

How many women understand the full implications of screening?

What do women believe?

Women generally exaggerate the benefits and are unaware of the harms of screening.[4–6] The authors of a study of American and European women[4] raised doubts about informed consent procedures since 68% believed screening reduced their risk of contracting breast cancer, 62% that screening at least halved mortality, and 75% that 10 years of screening saved 10 of 1000 participants, which is 10 times the most optimistic estimates.[7–9] Other studies have shown that only 8% of women were aware that participation has the potential to harm healthy women,[5] that 15% believe their lifetime risk of contracting the disease is more than 50% (an overestimate of about 5 times),[6] and that one third think screening detects more than 95% of breast cancers.[w5]

Assessment of invitations

We collected invitations to mammography screening from Australia, Canada, Denmark, New Zealand, Norway, Sweden, and the United Kingdom. These countries all have publicly funded screening programmes that are nationally or regionally coordinated and use languages we can read. We requested letters that invited women for the first time including any enclosed pamphlets, letters to non-responders, and invitations to subsequent screening rounds. We focused on the initial invitation, which is most commonly sent out as women turn 50 years of age. We contacted organising units by email, telephone, or post and made three requests in case of non-response.

References w1-w11 and details of invitations are on bmj.com

Material was collected between October 2004 and February 2005.

We recorded whether a date of appointment was issued in the invitation, whether a reminder letter or other means of contact was used for non-responders, whether suggestive headlines or appeals for participation were used, and whether regular breast self examinations, clinical breast examinations, or both were recommended, as an additional check on whether the information was evidence based.[10] We evaluated the invitations independently and settled any discrepancies by discussion. We used the same checklist of 17 information items on benefits and harms as in our previous study of websites,[2] most of which have been used in other studies of information materials.[1 w1]

We identified 51 coordinating units and 33 (65%) responded. The unit in southern Australia declined to supply a sample as it was revising its invitation, and Nova Scotia, Canada, did not issue invitations but used public advertising. Thus, we obtained samples from 31 areas, including all seven countries. For Norway and the UK, we evaluated a national sample letter. In the UK, this letter may be modified locally, but the accompanying pamphlet, which contains the bulk of the information, is the same. The Norwegian letter and pamphlet are used nationwide. The response rate was lower for Sweden than for other countries (9/22 regions), but this is unlikely to have influenced our overall findings because the information varied little among those that responded.

How are women invited?

Since the wording and contents differed little within each country, our main emphasis is on the national results (see table on bmj.com). Twenty one invitations (68%), at least one from each country, gave an appointment date, but in New Zealand women receive a letter only after registering with the programme. Reminder letters were used in 18 of 31 areas (58%) but not in Sweden or the UK. In New Zealand, women who do not attend their appointment are telephoned, and in Western Australia and some areas of New Zealand, a letter is sent to general practitioners informing them about non-responders and asking them to discuss this with the woman at the next consultation. These national differences limit the applicability of our results to countries we did not include, in particular countries where screening is a private enterprise.

What do the invitations contain?

The invitations included a median of 2 of the 17 possible information items in our survey, ranging from none in Sweden to six in New Zealand. A pamphlet was included with 20 invitations (65%), but only in one of nine invitations in Sweden.

Thirty invitations (97%) mentioned the main benefit of screening, a reduction in breast cancer mortality, but only seven (three countries) gave the size of the benefit and they all described it as a relative risk reduction rather than an absolute risk reduction or the number needed to screen. The effect of screening on total mortality was not mentioned. In contrast, no invitation mentioned the major harm of screening, overdiagnosis and subsequent overtreatment.

Six invitations (five countries) argued that screening leads to less invasive surgery and four additional invitations (one additional country) that it leads to simpler treatment. None of the invitations noted the uncertainties related to treatment of carcinoma in situ or the increased use of surgery and radiotherapy arising from overdiagnosis.

The most commonly mentioned harm was pain associated with the procedure (15 invitations (48%), six countries), but it was downplayed in eight—for example, "Any discomfort should only last a few seconds" (BreastScreen Western Australia). The lifetime risk of developing breast cancer was noted in 10 invitations (32%, six countries) and estimates varied from 1 in 9 to 1 in 13.

Recall rates for further examinations appeared in six invitations (19%), but as the risk in each screening round, not the accumulated risk. After 10 screens, the accumulated risk of recall is about 50% for American women[11] and about half as much for European women.[12] About a quarter of these women will have a biopsy or fine needle aspiration.[9] A false positive result can have a profound psychological effect on women and their families because it raises the suspicion of a potentially life threatening disease.[13]

Seven invitations mentioned screening sensitivity, but five were misleading. For example, the Manitoba pamphlet states that "about 1 out of every 10 breast cancers cannot be seen on a mammogram," a 90% detection rate. This obscures the fact that many, indeed the most dangerous, cancers are detected in the intervals between screening rounds.[3] Interval cancer rates of up to 50% are deemed acceptable with biennial screening according to European guidelines.[w4] Neither specificity nor positive predictive value was mentioned.

Fifteen invitations (48%) recommended regular breast self examination, clinical breast examinations, or both. This is despite evidence that self examination leads to a doubling in biopsy of benign growths and probably has no mortality benefit[10] and the lack of evidence for an effect of clinical breast examinations.[3 9]

Appeals for participation appeared in only one of nine letters in Sweden, but in 17 of the remaining 22—for example, "We strongly recommend that you use this free service" (Northern Territory, Australia). Seven reminder letters had stronger pleas than the first letter (box 1).

Nineteen pamphlets (95%) had suggestive headlines, such as, "Have a screening mammogram, it may save your life" (Western Australia) and "Why is having a breast screen a good idea?" (New South Wales, Australia).

Problems with current practice

Although it is good news that the invitations often included an information pamphlet, the focus on the benefits of screening is problematic. The benefits were framed positively, avoiding absolute risk reductions and number needed to treat, which are easier to understand and provide more realistic expectations.[14] The reduction in breast cancer mortality was given as 25-30%, although recent systematic reviews have either doubted the effect[8] or suggested relative risk reductions of 15%-21%.[9 15] These estimates do not convey that they apply only to the period when women are screened and are not a reduction in lifetime risk.

Box 1: Excerpts from information material

Invitational letters

"We have reserved a time ... If the time is very inconvenient, we ask you to contact the mammography screening centre as soon as possible" [our translation]—Funen, Denmark

"During the past two years, over 340 000 Queensland women have benefited from taking part in the BreastScreen Queensland Programme" [this refers to the number of participants; less than 0.1% of those would have benefited]

"You can take a positive step to decrease your own risk, and help us achieve our goal, by deciding to take part" [clear conflict of interest]—Northern Territory, Canada

"I am writing to personally invite you" [inappropriate use of familiarity]—Vancouver

Letters to non-attenders

"I do not wish to participate in the examination due to the following reason: ____" (return slip) [Is it acceptable to demand a reason for declining participation in something the woman didn't ask for?]—Funen, Denmark

"If you would like to avoid participation, we ask you to fill out a form. You obtain this form by calling the breast-diagnostic centre" [our translation; some work is required to avoid further invitations]—Norway

"Some of these studies have been going on for 30 years and none have found any serious side effects from the mammography" [this is wrong because overdiagnosis is a serious harm]—New South Wales

"I am concerned that you have not yet responded to our recent invitation for a screening mammogram (breast x ray) … Every year in NSW about 3000 women develop breast cancer and about 900 die from the disease" [Paternalistic, and tells non-attenders that they are behaving irresponsibly]

Pamphlets

"There has been a 26% increase in breast cancer cases in the last ten years" [scaring and misleading—this is the level of overdiagnosis expected with screening over the 10 years this programme had been operating]—Ontario

"Research has shown that regular screening mammograms can lower deaths in women 50 to 69 years of age by 1/3" [the risk of dying (total mortality) is reduced by 0.1% at most]—Manitoba

"The benefits of screening far outweigh the risks of any harm from the breastscreen" [subjective statement; this judgment should be left to the women]—Queensland

The most important harms, overdiagnosis and overtreatment, were not mentioned and other important harms were often either omitted or downplayed. The estimated level of overdiagnosis, 30% in the randomised trials,[8] is supported by large epidemiological studies that have suggested 40-60%.[16–20] Carcinoma in situ is a special case as it is rarely detected without screening and represents about 20% of all screen detected cancers.[3] Little is known about its natural course, but autopsy studies indicate that many lesions do not progress.[21] Because it is impossible to tell which lesions will become invasive, all are treated, often with mastectomy and radiotherapy.[8 9]

Overdiagnosis led to screening programmes for neuroblastoma in children being stopped[w6] and is a main reason why screening for prostate and lung cancer is generally discouraged.[22] Very few women are aware that screening can detect non-progressive cancer,[23] and probably even fewer know that invasive cancer can sometimes regress spontaneously.[w7] Many will falsely believe their lives have been saved by screening, when in fact they have only been physically and psychologically harmed.

Participation rates increase when there is a pre-assigned date of appointment,[3 24] but we find this approach problematic as it bypasses the informed consent step and gives the impression that participation is a public duty. Information material should convey the message that a decision not to attend mammography screening can be based on sound reasoning and is not irresponsible, as is currently believed by about 75% of 55 year old Americans.[5]

Fear of cancer seems to increase participation in breast cancer screening,[w8] and the frequent mention of lifetime risk of developing breast cancer in the information could scare some women to participate without considering the harms, especially as these were so rarely mentioned.

More comprehensive information will lead to more women declining to be screened.[w9] Uptake rates in Sweden are high, 78-84%,[3] which may be related to the fact that the invitations contain little information apart from a date of appointment and explanations of practical matters such as transport and payment of a small fee.

Implications

Informed consent cannot be achieved solely through information in invitations. It is a process that should include a discussion with a general practitioner, as preferred by 88% of Swiss women.[w10] It is not reasonable to assume that participants have been adequately

Box 2: Key elements in information leaflets

Main benefits and harms, assuming a 15% reduction in breast cancer mortality and overdiagnosis of 30%

If 2000 women are screened regularly for 10 years:

- 1 woman will avoid dying from breast cancer
- 10 healthy women, who would not have been diagnosed without screening, will have breast cancer diagnosed and be treated unnecessarily; 4 of these will have a breast removed, 6 will receive breast conserving surgery, and most will receive radiotherapy
- 1800 will be alive after 10 years; without screening 1799 will be alive.[8]

Other main points

Of 2000 women (in Europe) who participate in 10 rounds of screening

- 500 will be recalled for additional investigations because cancer is suspected; about 125 will have a biopsy[9]
- 200 will experience psychological distress for several months related to a false positive finding[8]

Screening can provide false reassurance. Up to 50% of cancers among women in screening programmes are detected between two screening rounds,[w4] and these interval cancers are the most dangerous

Mammography is painful for about a third of women[w11]

Summary points

- Conflict of interest exists for publicly funded screening since organisers want a high uptake
- No invitations contain information about the major harms of screening
- Most invitations use pre-specified appointments and persuasive wording
- The information sent to women needs to be more balanced
- Harms and benefits should be presented in more easily understandable ways

informed about important harms through other sources. We believe that the information included with invitations should be more balanced, using absolute numbers to describe the likelihood of benefits and harms,[19] and applying to the same time span if possible (box 2). Furthermore, we suggest that the responsibility for the programmes should be separated from the responsibility for the information material and that consumer groups be involved in the process of developing balanced information material.

We thank those who supplied material, Hazel Thornton for contact details for the UK screening programme, and Kay Dickersin for comments on our manuscript.

Contributors and sources: We have previously evaluated information on mammography screening on websites and PCG was involved in a systematic review of the breast screening trials, which questioned the value of screening. Both authors contributed to conceiving the project. The draft protocol was written by KJJ and revised by PCG. KJJ collected the invitations and wrote the first manuscript. Both authors extracted data and contributed to the interpretation and final manuscript. Both are guarantors.

Competing interests: None declared.

1 Slaytor EK, Ward JE. How risks of breast cancer and benefits of screening are communicated to women: analysis of 58 pamphlets. *BMJ* 1998;317:263-4.
2 Jørgensen KJ, Gøtzsche PC. Presentation on websites of possible benefits and harms from screening for breast cancer: cross sectional study. *BMJ* 2004;328:148.
3 Vainio H, Bianchini F. *IARC handbooks of cancer prevention. Vol 7: breast cancer screening*. Lyons: IARC Press, 2002.
4 Domenighetti G, D'Avanzo B, Egger M, Berrino F, Perneger T, Mosconi P, et al. Women's perception of the benefits of mammography screening: population-based survey in four countries. *Int J Epidemiol* 2003;32:816-21.
5 Schwartz LM, Woloshin S, Fowler FJ Jr, Welch HG. Enthusiasm for cancer screening in the United States. *JAMA* 2004;291:71-8.
6 Barratt AL, Cockburn J, Redman S, Paul C, Perkins J. Mammographic screening: results from the 1996 national breast health survey. *Med J Aust* 1997;167:521-4.
7 Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993;341:973-8.
8 Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001;358:1340-2. http://image.thelancet.com/extras/fullreport.pdf (accessed 8 Feb 2006).
9 Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;137:347-60.
10 Kösters JP, Gøtzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. *Cochrane Database Syst Rev* 2003;(2):CD003373.
11 Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338:1089-96.
12 Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, et al. Comparison of screening mammography in the United States and the United Kingdom. *JAMA* 2003;290:2129-37.
13 Aro AR, Pilvikki Absetz S, van Elderen TM, van der Ploeg E, van der Kamp LJ. False-positive findings in mammography screening induces short-term distress—breast cancer-specific concern prevails longer. *Eur J Cancer* 2000;36:1089-97.
14 Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *BMJ* 1995;311:1056-9.
15 Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002;359:909-19.
16 Gøtzsche PC. On the benefits and harms of screening for breast cancer. *Int J Epidemiol* 2004;33:56-64.
17 Zahl PH, Strand BH, Mæhlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004;328:921-4.
18 Douek M, Baum M. Mass breast screening: Is there a hidden cost? *Br J Surg* 2003;90:44-5.
19 Barratt A, Howard K, Irwig L, Salkeld G, Houssami N. Model of outcomes of screening mammography: information to support informed choices. *BMJ* 2005;330:936-8.
20 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.
21 Welch HG, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? *Ann Intern Med* 1997;127:1023-8.
22 Welch HG. *Should I be tested for cancer? Maybe not and here's why*. Los Angeles: University of California Press, 2004.
23 Schwartz LM, Woloshin S, Sox HC, Fischhoff B, Welch HG. US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *BMJ* 2000;320:1635-40.
24 Bonfill X, Marzo M, Pladevall M, Marti J, Emparanza JI. Strategies for increasing women participation in community breast cancer screening. *Cochrane Database Syst Rev* 2001;(1):CD002943.

(Accepted 1 December 2005)

Hoax bomb calls and psychiatry

Nowadays in Britain, it is not uncommon to read or hear a news headline about a person "arrested and charged in connection with a hoax telephone call warning of a bomb." Such hoaxers are often labelled insensitive, uncaring, or psychopaths, but a patient I recently saw in my clinic has made me see this phenomenon differently.

She started by saying how anxious she was to be in the city centre, given recent bomb alerts. I empathised with her and shared my not dissimilar concerns. She volunteered several possible explanations for the London bombings of July 2005—the Iraq war, anti-Western sentiments, Islamic fundamentalism, and so on. She added that her thyroid gland was swollen, that she was emitting a bile-like substance from her "waste pipes," and that she was in the process of designing a spacecraft for the salvation of humanity. A clear picture of psychosis emerged.

She continued by saying that she had seen a suspicious car abandoned in the hospital car park. It was red and very clean, which had led her to believe it belonged to a suicide bomber, as "they wash their blood and conscience off, don't they? So, it's always clean. Maybe, I should have called the police, but I didn't."

Some patients with psychotic disorders may incorporate recent public events into their existing delusional belief systems and occasionally even act out these delusions. My patient, among other delusions, held bizarre beliefs about suicide bombers (triggered by recent events) and nearly acted on this delusion by calling the police. We should consider such a possibility before blaming all people who make hoax calls. Their behaviour may be a manifestation of underlying psychopathology. Perhaps, all such calls should lead to the perpetrator being required to submit to mental health assessment.

Sanju George *specialist registrar in psychiatry, Queen Elizabeth Psychiatric Hospital, Edgbaston, Birmingham (sanju.george@talk21.com)*

BMJ

RESEARCH

Breast cancer mortality in organised mammography screening in Denmark: comparative study

Karsten Juhl Jørgensen, researcher,[1] Per-Henrik Zahl, senior researcher,[2] Peter C Gøtzsche, professor[1]

[1]The Nordic Cochrane Centre, Rigshospitalet, University of Copenhagen, Denmark

[2]Norwegian Institute of Public Health, Oslo, Norway

Correspondence to: K J Jørgensen
kj@cochrane.dk

Cite this as: *BMJ* 2010;340:c1241
doi:10.1136/bmj.c1241

ABSTRACT

Objective To determine whether the previously observed 25% reduction in breast cancer mortality in Copenhagen following the introduction of mammography screening was indeed due to screening, by using an additional screening region and five years additional follow-up.

Design We used Poisson regression analyses adjusted for changes in age distribution to compare the annual percentage change in breast cancer mortality in areas where screening was used with the change in areas where it was not used during 10 years before screening was introduced and for 10 years after screening was in practice (starting five years after introduction of screening).

Setting Copenhagen, where mammography screening started in 1991, and Funen county, where screening was introduced in 1993. The rest of Denmark (about 80% of the population) served as an unscreened control group.

Participants All Danish women recorded in the Cause of Death Register and Statistics Denmark for 1971-2006.

Main outcome measure Annual percentage change in breast cancer mortality in regions offering mammography screening and those not offering screening.

Results In women who could benefit from screening (ages 55-74 years), we found a mortality decline of 1% per year in the screening areas (relative risk (RR) 0.99, 95% confidence interval (CI) 0.96 to 1.01) during the 10 year period when screening could have had an effect (1997-2006). In women of the same age in the non-screening areas, there was a decline of 2% in mortality per year (RR 0.98, 95% CI 0.97 to 0.99) in the same 10 year period. In women who were too young to benefit from screening (ages 35-55 years), breast cancer mortality during 1997-2006 declined 5% per year (RR 0.95, CI 0.92 to 0.98) in the screened areas and 6% per year (RR 0.94, CI 0.92 to 0.95) in the non-screened areas. For the older age groups (75-84 years), there was little change in breast cancer mortality over time in both screened and non-screened areas. Trends were less clear during the 10 year period before screening was introduced, with a possible increase in mortality in women aged less than 75 years in the non-screened regions.

Conclusions We were unable to find an effect of the Danish screening programme on breast cancer mortality. The reductions in breast cancer mortality we observed in screening regions were similar or less than those in non-screened areas and in age groups too young to benefit from screening, and are more likely explained by changes in risk factors and improved treatment than by screening mammography.

INTRODUCTION

Comprehensive systematic reviews of randomised trials of mammography screening have estimated that mammography reduces breast cancer mortality by 15-16%.[1 2] The trials in these reviews were carried out decades ago, however, and publicly available screening programmes could yield a different effect from that in the trials because of differences in the qualifications of the staff, type of equipment, and uptake rates. Furthermore, there have been advances in treatment since the trials were completed and "breast awareness" has increased. It is therefore important to evaluate continuously the effect of public mammography screening programmes to ensure that they live up to expectations.

Denmark is uniquely suited for observational studies of mammography screening because the country has had a period of 17 years where only about 20% of the population has been offered screening; that is, there is a concomitant non-screened control group. There are, however, substantial problems in using observational studies to estimate the effect of screening.[3] For example, a decline in breast cancer mortality following the introduction of screening would not necessarily be caused by screening.

A cohort study from 2005 by Olsen et al compared Copenhagen, where screening was introduced in 1991, with non-screened areas in Denmark and reported a 25% reduction in breast cancer mortality that was attributed to screening.[4] However, there are three important concerns about this result.[5-7] Firstly, the full mortality reduction appeared three years after screening started and did not increase in the remaining observation period.[4] The mechanism of screening is to advance the time of diagnosis; therefore, an effect is not expected to appear in the first few years after its introduction but is expected to emerge after about five years and increase with further follow-up.[3] That the full effect appeared after only three years suggests that factors other than screening are the cause of the mortality reduction observed.

Secondly, the study included only Copenhagen, although the county of Funen introduced screening in 1993, has a population of a similar size, and has a higher proportion of women who repeatedly attend screening rounds (76% *v* 53%).[8] One of the authors of the paper by Olsen et al heads the Funen screening programme and should have had access to detailed data.

Thirdly, the study did not describe breast cancer mortality rates in women who were too young or too old to have benefited from screening. An absence of similar reductions in breast cancer mortality among these women would have strengthened the study's conclusions.

We hypothesised that if the reduction in breast cancer mortality that Olsen et al observed in Copenhagen was due to screening, a similar reduction should have occurred in Funen but not in the non-screened areas or in age groups outside those that could potentially have benefited.

METHODS

We retrieved data on female breast cancer mortality during 1971-2006 from the Cause of Death Register through the National Board of Health. Compared with Olsen et al,[4] we had access to data from five additional years. The numbers of breast cancer deaths were listed for each year, administrative region, and five year age group. The corresponding female population statistics were obtained from Statistics Denmark.[9]

Organised mammography screening of women aged 50-69 years began on 1 April 1991 in Copenhagen municipality, 1 November 1993 in Funen county, and 1 June 1994 in Frederiksberg municipality.[10] The Frederiksberg programme, which comprised only about 10 000 women, was incorporated into the Copenhagen programme on 1 January 1997.[10] The Copenhagen and Funen programmes both include about 50 000 women.

We divided the data into two regions. Copenhagen (including Frederiksberg municipality) and Funen county were considered together as "screening areas" to reduce the effect of random fluctuations. These screening areas were compared with the "non-screened areas," comprising the rest of Denmark (about 80% of the population).

The mortality data for both screening and non-screening areas were divided into three age bands. The 55-74 years band was composed of the women most likely to have benefited from a programme targeted at women aged 50-69 years, such as that in Denmark. Most women who die aged 50-55 years would have had their breast cancers detected before they were invited to screening. On the other hand, the majority of women aged 70-74 years when they died would have been diagnosed and offered screening when they were younger, and by six years after organised screening began, all women aged 70-74 years would have been previously offered screening.

In contrast, breast cancer mortality in women aged 35-54 years and 75-84 years would largely be unaffected by screening, although by the end of the observation period some women aged 75-84 years could have benefited through detection of slow growing cancers. These age groups also serve as control groups.

We first defined the period when screening could have had an effect. In randomised trials of mammography screening, an effect began to emerge about five years after screening was introduced,[3] which is 1996 in Copenhagen and 1998 in Funen. We used 1997 as a compromise start date in our combined analysis of the two regions. This provided a 10 year observation period (1997-2006) when screening could have had an effect.

For comparison, we used a 10 year period that ended when screening was introduced (1982-1991), but we provide mortality graphs back to 1971 for completeness. The last year before organised screening was introduced was 1990 in Copenhagen and 1992 in Funen, so we used 1991 as a compromise.

We used Poisson regression analyses to quantify changes in mortality trends, comparing regions and age groups with and without organised mammography

Table 1 | Woman years of observation, number of breast cancer deaths, and average rates per 100 000 women in screened and non-screened areas

	Screened areas			Non-screened areas		
	1982-91	1992-96	1997-2006	1982-91	1992-96	1997-2006
Number of woman years						
35-54 years	1 174 997	640 700	1 338 266	5 509 889	3 005 488	6 233 027
55-74 years	1 218 157	492 081	957 797	3 925 135	2 002 687	4 471 369
75-84 years	478 800	223 441	370 101	1 142 623	625 693	1 312 095
Number of breast cancer deaths						
35-54 years	457	257	390	1961	1055	1657
55-74 years	1478	658	980	4281	2326	4739
75-84 years	937	429	722	2003	1152	2352
Average number of breast cancer deaths per 100 000 women						
35-54 years	39	40	29	36	35	27
55-74 years	121	134	102	109	116	106
75-84 years	196	192	195	175	184	179

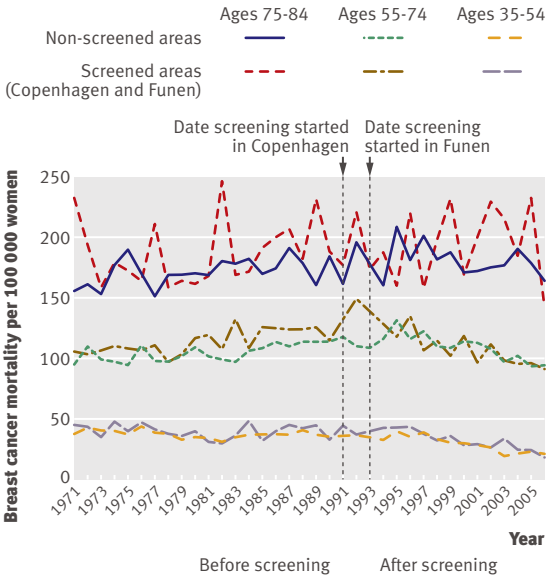

Fig 1 | Unadjusted breast cancer mortality rates for screened and non-screened areas in Denmark

screening and correcting for changes in age distribution. Statistical analyses were made using Egret version 2.0.3 (Cytel, Inc, Cambridge, MA) and graphs were made in Microsoft Excel 2000.

RESULTS

The number of woman years and the number of deaths attributed to breast cancer in the 36 year observation period in the screened areas and in the non-screened areas are presented in table 1.

Breast cancer mortality among women who could benefit from screening (ages 55-74 years)

In the 10 year period when screening could have had an effect, breast cancer mortality among women who could benefit from screening (ages 55-74 years) declined by 1% a year in the screened areas (relative risk (RR) 0.99; 95% confidence interval (CI) 0.96 to 1.01) and by 2% in the non-screened areas (RR 0.98, 95% CI 0.97 to 0.99; fig 1; table 2). Before screening was introduced, breast cancer mortality rates increased by 1% a year in the screened areas (RR 1.01, 95% CI 0.99 to 1.03) and by 2% a year in the non-screened areas (RR 1.02, 95% CI 1.01 to 1.03).

Breast cancer mortality among women too young to benefit from screening (ages 35-54 years)

In the most recent 10 year period, breast cancer mortality among women too young to benefit from screening (ages 35-54 years) declined by 5% a year in the screened areas (RR 0.95, 95% CI 0.92 to 0.98) and by 6% in the non-screened areas (RR 0.94, 95% CI 0.92 to 0.95; fig 1). Before screening was introduced, breast cancer mortality rates increased by 2% a year in the screened areas (RR 1.02, 95% CI 0.99 to 1.06) and were stable in the non-screened areas (RR 1.00, 95% CI 0.99 to 1.02; table 2).

Breast cancer mortality among women too old to benefit from screening (ages 75-84 years)

There were no significant changes in the breast cancer mortality trends among women too old to benefit from screening (ages 75-84 years), both in the screened areas and in the non-screened areas (fig 1; table 2).

DISCUSSION

Our findings contradict the conclusions of a recent observational study that reported a 25% reduction in breast cancer mortality in screening areas of Denmark (Copenhagen) compared with non-screened areas and attributed this drop to the effect of screening.[4] Our study included three screening areas in Denmark, non-screened age groups, and an additional five years of follow-up, yet we were not able to find an effect of the Danish screening programme on breast cancer mortality. We also note that in the age group too young to have benefited from screening, women experienced proportionately larger reductions in breast cancer mortality after screening was introduced than did those that could have benefited from screening.

The reduction in breast cancer mortality in Copenhagen started too early after screening was introduced for screening to be a plausible cause, although a decline within the first three years of screening formed the basis for the conclusion in the previous study,[4] and the reduction in breast cancer mortality in the second screened region, Funen, began before organised screening (see web extra). This suggests that causes other than screening were responsible for the changes in breast cancer mortality—for example, changes in risk factors and improvements in treatment may have occurred sooner in some areas than in others.

Strengths and weaknesses of our study

Opportunistic screening is rare in Denmark so could not be the reason for the reduction in breast cancer mortality in the non-screened areas. Only 2% of women aged 20-99 years and 3% of women aged 50-69 years had a mammogram outside organised screening in 2000, and some of the mammograms were diagnostic

Table 2 | Annual change in the relative risk of breast cancer death (with 95% confidence intervals) 10 years before screening was introduced and 10 years during which screening could have had an effect on breast cancer mortality

	Before screening (1982-91)	After screening (1997-2006)
Ages 35-54 years		
Screened areas	1.02 (0.99 to 1.06)	0.95 (0.92 to 0.98)
Non-screened areas	1.00 (0.99 to 1.02)	0.94 (0.92 to 0.95)
Ages 55-74 years		
Screened areas	1.01 (0.99 to 1.03)	0.99 (0.96 to 1.01)
Non-screened areas	1.02 (1.01 to 1.03)	0.98 (0.97 to 0.99)
Ages 75-84 years		
Screened areas	0.99 (0.97 to 1.02)	1.00 (0.98 to 1.03)
Non-screened areas	0.99 (0.98 to 1.02)	0.99 (0.98 to 1.02)

rather than screening mammograms.[11] Carcinoma in situ, which is detected almost entirely through screening, had a fairly constant incidence rate throughout the observation period in the non-screened areas, but the rates doubled in the screened regions when screening was introduced and have remained at that level.[12]

In the study by Olsen et al, deaths from breast cancers diagnosed before 1991, when screening was introduced in Copenhagen, were excluded from the analysis.[4] It may be reasonable to exclude breast cancers detected before screening was introduced because screening cannot affect their prognosis. Excluding such cancers is unlikely to have had an effect on our results, however, because the effect diminishes with time. Therefore, a difference in breast cancer mortality of 25% between screened and non-screened areas, if true, would be expected to be clearly apparent after 14-16 years with organised screening.

Furthermore, in 1986, before screening was introduced in the United Kingdom, the Forrest report on screening for breast cancer noted that population statistics could be used to see the mortality benefit from screening that was expected on the basis of results from randomised trials.[13] We did not see this expected effect in Denmark.

We compared open cohorts because our data did not allow identification of individual women. We could not therefore take account of migration between screened and non-screened regions. However, mobility in the screened age groups is limited in Denmark. We also note that our open cohort design cannot be used to explain why breast cancer mortality declined in the non-screened areas at a similar rate as in the screened areas, and that the decline was larger in women who were too young to have benefited from screening. We consider it unlikely that there were regional differences in the use of hormone replacement therapy that could explain our findings because Denmark is a homogeneous country and the screened and the non-screened areas had a similar proportion of cities and rural areas.

Comparison with previous studies

Olsen et al reported the results of a complicated statistical model, and the choice of model can have a substantial impact on the results. Some of the same authors have more recently reported results from several models, one of which showed an increase in breast cancer mortality in Copenhagen among screened women relative to women in the non-screened areas.[14] The authors asserted that the model they published originally, which gave the most favourable result for screening, should be preferred, but did not explain why. None of the models was validated, and it is not clear whether the preferred model was selected after other models had been tried first. Our study does not use a complicated statistical model; instead we present the raw data and simple analyses.

Olsen et al assumed that women in the screened areas and those in the non-screened areas received identical care because national guidelines for breast cancer treatment have been in use in Denmark since 1977.[4] Guidelines, however, may not be used to the letter in clinical practice. In January 2007, for example, several women in one of the non-screened regions were compensated financially for having received treatment that did not live up to "best specialist standards."[15 16] If the standard of care were lower in the non-screened areas, it would only strengthen our findings that screening cannot be responsible for the declines in breast cancer mortality we observed.

All breast cancer treatment in Copenhagen takes place at university hospitals that provide specialised care and have a high volume of patients, whereas care is more variable in Funen, and there has been considerable pressure to centralise breast cancer treatment in other regions.[15] The reduction in breast cancer mortality in Copenhagen started earlier than expected after screening was introduced and the reduction in breast cancer mortality in Funen began before organised screening started (see web extra); therefore the mortality difference between Copenhagen and non-screened regions found in the previous study might partly reflect better organisation of treatment and earlier implementation of improvements in therapy (for example, use of tamoxifen).

Olsen et al calculated that the reduction in breast cancer mortality among those who actually attended screening was 37%.[4] However, there was no relevant control group to compare those who actually attended screening with, because it is impossible to know which women in the non-screened areas would have attended screening had it been offered. Such results are invalid because of the "healthy screenee effect." People who attend screening are more healthy in general than those who choose not to participate, and have been described as "healthy, well educated, affluent, physically fit, fruit and vegetable eating, non-smokers with long lived parents."[17]

It is common for mortality estimates to be adjusted for attendance to mammography screening, but this measure is equivalent to preferring a per protocol analysis over an intention to treat analysis. By comparison, one does not adjust for compliance in drug trials because such adjustments are bias prone. For example, in a trial that failed to find an effect of clofibrate on cardiac mortality, the authors reported a large effect among those who took at least 80% of the drug (P=0.0001) but found a similar "effect" among compliers in the placebo group ($P=5\times10^{-16}$).[18]

The handbook on breast cancer screening from the International Agency for Research on Cancer advises that observational studies should not be regarded as providing evidence of an effect of screening.[3] We agree that positive evidence for an effect should come from randomised trials. The effect of screening is equivocal and has been estimated to be only about 15-16% in the two most comprehensive systematic reviews of randomised trials.[1 2] Small effect sizes render observational studies particularly problematic, but observational studies are useful for monitoring the effect in clinical practice.[12] Contrary to expectations, a study

in Europe found that declines in breast cancer mortality were of a similar magnitude in countries not offering screening as in those offering screening, with the greatest declines among women who were too young to be offered screening, which is similar to our findings.[19]

Many studies have used population statistics to estimate the effect of mammography screening programmes.[20-23] Like the study by Olsen et al, most of them include only women in the age group that could benefit from screening and do not compare trends in breast cancer mortality with those in an unscreened group, thus disregarding the effects of other important factors that could change in the screening period (such as treatment).

These studies also often claim to have compensated for lead time and length bias, with highly variable assumptions of their size. This is not surprising, as no one knows exactly how large lead time and length bias for breast cancer screening are. Such shortcomings question the conclusions in these studies.

Comparison with other countries

Results from other countries[19] support our findings. In the UK, where screening started in 1988, the decline in breast cancer mortality from 1989 until 2007 was 41% in women aged 40-49 years, who were not invited to screening, 41% in women aged 50-64 years, who were invited to screening from 1988, and 38% in women aged 65-69 years, who were invited from 2002 (fig 2).[24] Furthermore, the drop in breast cancer mortality in the relevant age group began before the screening programme started, and was largest in the age group that was too young to be invited (40-49 years) if the whole observation period is considered (1971-2007).[25]

In the United States, a particularly pronounced trend shift in breast cancer mortality has been observed, with a substantial decline since the middle of the 1980s.[27] However, the starting point for the decline is probably a random high, and there have been large changes in the use of surgery and chemotherapy that might affect the drop.[28] Groups of epidemiologists analysing these data independently arrived at a median estimate of the effect of screening of only 15%.[27]

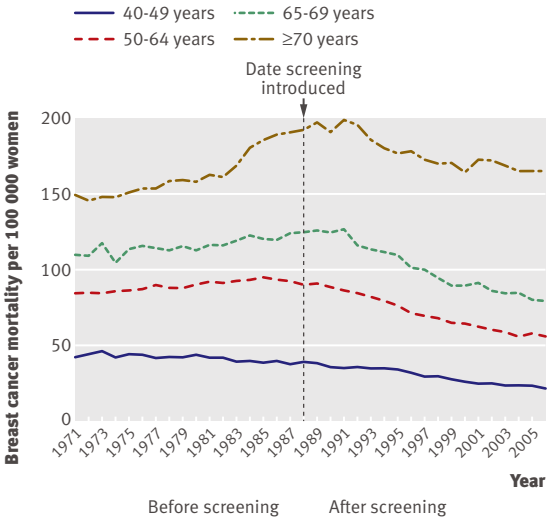

Fig 2 | Age adjusted breast cancer mortality rates in the United Kingdom for screened and non-screened age groups. Data from Cancer Research UK[26]

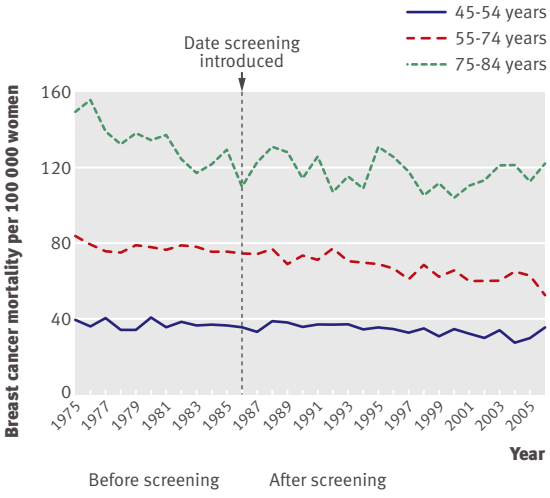

Fig 3 | Unadjusted breast cancer mortality rates in Sweden for screened and non-screened age groups. Data from Statistics Sweden (http://www.scb.se/default____2154.aspx)

In countries with a less pronounced trend shift following the introduction of mammography screening, or no shift at all,[19] such analyses would likely fail to find an overall positive effect of screening on breast cancer mortality. We note, for example, that in Sweden, where screening started in 1986 and where uptake rates are very high, the mortality curves for the relevant age groups (45-55 years and 55-74 years) show a constant decline through the decades (fig 3). This decline started before screening was introduced, which does not suggest a screening effect.

Furthermore, the decline in the youngest age group (45-54 years, some of whom are offered screening in some areas in Sweden) is similar to that in the age group that would be expected to benefit the most from screening (55-74 years, who stand to benefit from screening in all counties). Somewhat surprisingly, there was a pronounced decline in breast cancer mortality in the oldest age group (75-84 years) that stopped when screening was introduced (fig 3). This is contrary to what would be expected and is unlikely to have anything to do with screening.

Conclusions

We were unable to find an effect of the Danish screening programme on breast cancer mortality. The reductions in breast cancer mortality we observed in screened regions were similar or larger in non-screened regions and in age groups younger than that screened. The mortality reduction is therefore more likely to be explained by changes in risk factors and by improved treatment than by screening mammography. Our results are similar to what has been observed in other countries with nationally organised programmes. We believe it is

WHAT IS ALREADY KNOWN ON THIS TOPIC

A Danish study has estimated that breast cancer mortality in Copenhagen has fallen by 25% since 1991 following the introduction of screening mammography

However, the full mortality reduction appeared three years after screening started (that is, several years earlier than expected), the study did not describe breast cancer mortality rates in women who were too young or too old to have benefited from screening, and only one of two available screening regions was included

If the reduction in breast cancer mortality in Copenhagen was due to screening, a similar reduction should be seen in the other region of Denmark that used screening, but not in non-screened areas or in age groups other than those that could potentially have benefited

WHAT THIS STUDY ADDS

Among women who could benefit from screening (ages 55-74 years), there was a similar or larger decline in breast cancer mortality among women in areas that did not use screening than in those that did

The reductions in breast cancer mortality among women too young to benefit from screening (ages 35-54 years) were much larger than those among women in the screened age groups

The reductions in breast cancer mortality we observed are more likely explained by changes in risk factors and by improved treatment than by screening mammography

time to question whether screening has delivered the promised effect on breast cancer mortality.

Contributors: KJJ and PCG conceived the project. KJJ collected data and wrote the first draft. PHZ performed the statistical analyses. PHZ and PCG revised the manuscript.
Funding: No funding was received for this research.
Conflicts of interest: All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare (1) No financial support for the submitted work from anyone other than their employer; (2) No financial relationships with commercial entities that might have an interest in the submitted work; (3) No spouses, partners, or children with relationships with commercial entities that might have an interest in the submitted work; (4) No non-financial interests that may be relevant to the submitted work.
Ethical approval: No approval from an ethical committee or informed consent was needed.
Data sharing: The mortality data used in our calculations and graphs are available from the corresponding author at kj@cochrane.dk.

1 Humphrey LL, Helfand M, Chen BK, Woolf SH. Breast cancer screening: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med* 2002;137:347-60.
2 Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2006;4:CD001877.
3 Vainio H, Bianchini F. IARC handbooks of cancer prevention vol 7: breast cancer screening. IARC Press, 2002.
4 Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB, et al. Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. *BMJ* 2005;330:220-4.
5 Gøtzsche PC, Thornton H, Jørgensen KJ. Reduction in mortality from breast cancer: presentation of benefits and harms needs to be balanced. *BMJ* 2005;330:1024.
6 Zahl PH, Mæhlen J. Reduction in mortality from breast cancer: decrease with screening was marked in younger age group. *BMJ* 2005;330:1024.
7 Lynge E, Olsen AH. Reduction in mortality from breast cancer: authors reply. *BMJ* 2005;330:1025.
8 Euler-Chelpin M, Olsen AH, Njor S, Vejborg I, Schwartz W, Lynge E. Women's patterns of participation in mammography screening in Denmark. *Eur J Epidemiol* 2006;21:203-9.
9 Statistics Denmark. www.statistikbanken.dk. (accessed Sept 2008)
10 Danish National Board of Health. *Tidlig opsporing og behandling af brystkræft* [Early detection and treatment of breast cancer: status report]. Copenhagen, 1997.
11 Jensen A, Olsen AH, von Euler-Chelpin M, Njor SH, Vejborg I, Lynge E. Do nonattenders in mammography screening programmes seek mammography elsewhere? *Int J Cancer* 2005;113:464-70.
12 Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009;9:36.
13 Forrest P. *Breast Cancer Screening. Report to the Health Ministers of England, Wales, Scotland & Northern Ireland*. Department of Health and Social Science, 1986.
14 Olsen AH, Njor SH, Lynge E. Estimating the benefits of mammography screening. The impact of study design. *Epidemiology* 2007;18:487-92.
15 Pedersen FS. [Cancer victims to be compensated for unnecessary breast surgery]. *Berlingske Tidende*, 25 January 2007.
16 [Compensation for removed breasts]. *Jylland-Posten*, 25 January 2007.
17 Raffle AE, Gray M. Screening: evidence and practice. Oxford University Press, 2007.
18 Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980;303:1038-41.
19 Levi F, Bosetti C, Lucchini F, Negri E, Vecchia CL. Monitoring the decrease in breast cancer mortality in Europe. *Eur J Cancer Prev* 2005;14:497-502.
20 Duffy S, Tabar L, Chen HH, Holmquist M, Yen MF, Abdsalah S, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties. *Cancer* 2002;95:458-69.
21 The Swedish organised service screening evaluation group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev* 2006;15:45-51.
22 Gabe R, Duffy S. Evaluation of service screening mammography in practice: the impact on breast cancer mortality. *Ann Oncol* 2005;16(suppl 2):ii153-62S.
23 Lawrence G, Wallis M, Allgood P, Nagtegaal ID, Warwick J, Cafferty FH, et al. Population estimates of survival in women with screen-detected and sympotomatic breast cancer taking account of lead time and length bias. *Breast Cancer Res Treat* 2009;116:179-85.
24 Mayor S. UK deaths from breast cancer fall to lowest figure for 40 years. *BMJ* 2009;338:b1710.
25 Jørgensen KJ, Brodersen J, Nielsen MN, Hartling OJ, Gøtzsche PC. A cause for celebration, and caution. *BMJ* 2009;338:1288.
26 Cancer Research UK. Breast cancer—UK mortality statistics. http://info.cancerresearchuk.org/cancerstats/types/breast/mortality/.
27 Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Cancer intervention and surveillance modelling network (CISNET). Effect of screening and adjuvant therapy on mortality from breast cancer. *New Engl J Med* 2005;353:1784-92.
28 Gøtzsche PC. Screening and breast cancer. *N Engl J Med* 2006;354:78. http://content.nejm.org/cgi/reprint/354/7/767.pdf.

Accepted: 25 November 2009

BMJ

RESEARCH

Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends

Karsten Juhl Jørgensen, researcher Peter C Gøtzsche, director

The Nordic Cochrane Centre, Rigshospitalet, Dept 3343, Blegdamsvej 9, DK-2100 Copenhagen, Denmark

Correspondence to: K J Jørgensen kj@cochrane.dk

Cite this as: *BMJ* 2009;339:b2587
doi:10.1136/bmj.b2587

ABSTRACT

Objective To estimate the extent of overdiagnosis (the detection of cancers that will not cause death or symptoms) in publicly organised screening programmes.

Design Systematic review of published trends in incidence of breast cancer before and after the introduction of mammography screening.

Data sources PubMed (April 2007), reference lists, and authors.

Review methods One author extracted data on incidence of breast cancer (including carcinoma in situ), population size, screening uptake, time periods, and age groups, which were checked independently by the other author. Linear regression was used to estimate trends in incidence before and after the introduction of screening and in older, previously screened women. Meta-analysis was used to estimate the extent of overdiagnosis.

Results Incidence data covering at least seven years before screening and seven years after screening had been fully implemented, and including both screened and non-screened age groups, were available from the United Kingdom; Manitoba, Canada; New South Wales, Australia; Sweden; and parts of Norway. The implementation phase with its prevalence peak was excluded and adjustment made for changing background incidence and compensatory drops in incidence among older, previously screened women. Overdiagnosis was estimated at 52% (95% confidence interval 46% to 58%). Data from three countries showed a drop in incidence as the women exceeded the age limit for screening, but the reduction was small and the estimate of overdiagnosis was compensated for in this review.

Conclusions The increase in incidence of breast cancer was closely related to the introduction of screening and little of this increase was compensated for by a drop in incidence of breast cancer in previously screened women. One in three breast cancers detected in a population offered organised screening is overdiagnosed.

INTRODUCTION

Screening for cancer may lead to earlier detection of lethal cancers but also detects harmless ones that will not cause death or symptoms. The detection of such cancers, which would not have been identified clinically in someone's remaining lifetime, is called overdiagnosis and can only be harmful to those who experience it.[1] As it is not possible to distinguish between lethal and harmless cancers, all detected cancers are treated. Overdiagnosis and overtreatment are therefore inevitable.[2]

It is well known that many cases of carcinoma in situ in the breast do not develop into potentially lethal invasive disease.[1] In contrast, many find it difficult to accept that screening for breast cancer also leads to overdiagnosis of invasive cancer. Harmless invasive cancer is common, however, even for lung cancer, with 30% overdiagnosis after long term follow-up of patients screened by radiography.[2] Autopsy studies have shown that invasive prostate cancer occurs in about 60% of men in their 60s, whereas the lifetime risk of dying from such cancer is only about 3%.[2] Autopsy studies have also found inconsequential breast cancer lesions. Thirty seven per cent of women aged 40-54 who died from causes other than breast cancer had lesions of invasive or non-invasive cancer at autopsy, and half were visible on radiography.[3 4]

Overdiagnosis can be measured precisely in a randomised trial with lifelong follow-up if people are assigned to a screening or control group for as long as screening would be offered in practice, which in most countries is 20 years. Overdiagnosis would be the difference in number of cancers detected during the lifetime of the two groups, provided the control group or age groups not targeted are not screened. In the absence of overdiagnosis the initial increase in cancers in the screened age groups would be fully compensated for by a similar decrease in cancers among older age groups no longer offered screening, as these cancers would already have been detected.

The extent of overdiagnosis and overtreatment as a result of mammography screening was first quantified in reviews of randomised trials.[5 6] The total number of mastectomies and lumpectomies increased by 31% and mastectomies by 20%.[6] As these trials did not have lifelong follow-up the extent of overdiagnosis could have been overestimated. Underestimation is also possible, however, as the randomised design was

maintained for only 4-9 years[6] and as opportunistic screening occurred in the control groups.[7]

Screening programmes differ from randomised trials. Radiologists outside a rigorous trial setting may be less well trained than those in the trial, and technical developments resulting in higher resolution images may also affect outcomes. The basic premise of an unchanged lifetime risk of breast cancer in the absence of overdiagnosis is, however, the same.

To estimate the extent of overdiagnosis in organised screening programmes we compared trends in breast cancer incidence before and after screening, taking account of changes in the background incidence and any compensatory drop in incidence of breast cancer among older, previously screened women. We combined our results in a meta-analysis.

METHODS

We included articles in any language with data on breast cancer incidence for both screened and older, non-screened age groups for at least seven years before screening and seven years after screening had been fully implemented, regardless of the time it took to implement screening. We reasoned that a long period after implementation was necessary to obtain an estimate of the trend in breast cancer incidence that was unaffected by the initial peak in prevalence when screening is introduced. Acquiring incidence data for age groups older than those screened allowed us to evaluate any compensatory declines in incidence among previously screened women.

When a country was described in several papers we selected the one with the most recent and best reported data as our core article, and we supplemented with other papers when relevant. When possible we also added data from the internet and supplied by authors. We did not search for articles published before 1990, as insufficient time would have elapsed after the initiation of screening.

Literature searches

Our searches in PubMed were developed iteratively and we tried several search strings. The final search, which identified all included articles, was: ((“Mammography”[MeSH] OR “Mass Screening”[MeSH]) AND ((“Breast Neoplasms/epidemiology”[MeSH]) OR (“Breast Neoplasms”[MeSH] AND incidence*[ti]))) OR (Breast cancer AND screening AND trend*[ti]) OR (Breast cancer AND screening AND overdiagnos*[ti]).

One author scanned titles and abstracts and retrieved the full text of potentially relevant articles for evaluation of eligibility, scanned the reference lists, and contacted authors. We compared the final search with an archive of all articles on breast cancer screening published in 2004, which we have used for another study,[8] and found that we had not missed any potentially relevant papers. None of the four authors we contacted told us of additional studies but three provided unpublished data or referred us to internet resources.[9] We did not find additional studies in the reference lists.

Data extraction

Both authors extracted data independently, with differences resolved by discussion. We extracted data on population size, screening uptake, length of time before and after the implementation of screening, and incidence of breast cancer for both screened and non-screened age groups. If data on carcinoma in situ were missing, we estimated overdiagnosis with these cases included, assuming that they would contribute 10% of the diagnoses in a population offered screening[10 11] —that is, we divided the incidence of invasive cancers by 0.9.

Selection of last prescreening year

The last prescreening year was usually the year before formal implementation of screening. If the levels of invasive breast cancer or carcinoma in situ appeared to increase abruptly in the years immediately before the introduction of screening, however, we excluded these years from estimates of trends before screening. Carcinoma in situ is rarely diagnosed without screening and such increases indicate opportunistic screening (screening outside the organised programme). Similarly, abruptly increased rates of invasive breast cancer immediately before formal implementation of screening likely indicate pilot programmes or extensive opportunistic screening.

Calculation of overdiagnosis in absence of compensatory drop

We used simple linear regression to estimate trends as we could not use Poisson regression because the denominators for the reported rates of breast cancer were not available. To compensate for changes in background incidence in the screened age group we carried out a linear regression analysis of the prescreening years and extended this regression line to the last observation year. We used the calculated value for this year to estimate what the expected incidence would have been in the absence of screening.

We did another linear regression analysis for the screened age group but used the observed incidence in that part of the screening period where the programme was fully implemented and past any prevalence peak. This was done to take account of annual fluctuations. The rate ratio between the result for the last observation year determined by linear regression and the expected incidence in that year (that is, the observed incidence in the last observation year divided by the expected incidence in the last observation year) constituted our estimate of overdiagnosis.

Calculation of overdiagnosis in presence of compensatory drop

In the age group that exceeded the age for screening, we studied whether the observed increase in the

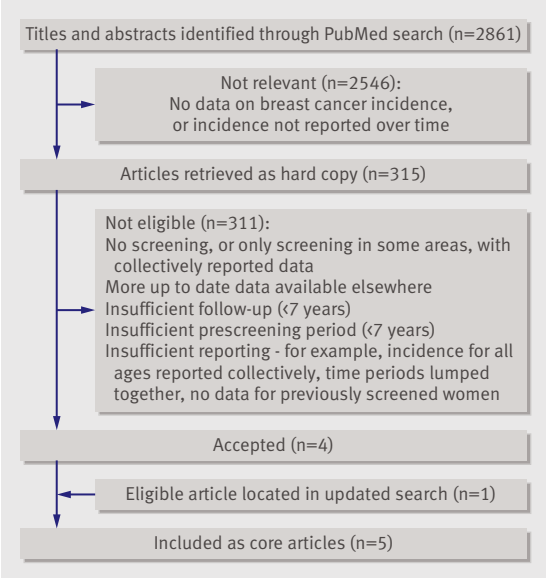

Fig 1 | **Selection of core articles**

incidence of breast cancer in the screening period was lower than the expected increase, in both cases using linear regression. If this was the case, we considered that the difference between the observed and expected incidence was due to a compensatory drop. We calculated the size of this drop as a rate ratio, as above, using the last observation year.

From this rate ratio we calculated the absolute number of breast cancer cases per 100 000 women that corresponded to the drop in the older age groups (X). Similarly, for the screened age groups we calculated the number of extra cases of breast cancer (those above the expected number) per 100 000 women that corresponded to the increase (Y). We compensated for the many more women in the younger, screened age group (A) than in the older age group of previously screened women (B) using official population statistics to calculate a correction factor C=A/B.

We calculated (Y×C−X)/(Y×C), which is the percentage of breast cancer cases uncompensated for of the total percentage increase in incidence among screened women. Overdiagnosis was then the observed percentage increase in incidence multiplied by the percentage of uncompensated for breast cancers (see Manitoba under Results for a numerical example).

Women too young to be screened

If available we used the group of women who were too young to be screened as a control to see if our extrapolated prescreening trend for the screened age group was a reasonable estimate of the background incidence, if screening had not been introduced. We did a linear regression analysis using the prescreening incidence, extrapolated the trend into the screening period, as for the other age groups, and compared with the observed incidence.

Meta-analysis

We combined the estimates using Comprehensive Meta Analysis version 2.2.046 (random effects model). As we estimated overdiagnosis using only the last observation year, our estimate has wider confidence intervals than if we had used several observation years. We used population sizes and age distributions obtained from internet sources[9] or as provided by the authors.

RESULTS

The PubMed search (May 2006) yielded 2861 titles, 2546 of which were not relevant (fig 1). The full text of the remaining 315 articles was evaluated for eligibility. Four were included as core articles and one was added when the search was updated in April 2007, presenting data from the United Kingdom; Manitoba,

Overview of individual estimates of overdiagnosis for invasive breast cancer, excluding cases of carcinoma in situ except for Manitoba, Canada

Variables	United Kingdom	Manitoba, Canada	New South Wales, Australia	Sweden	Norway (AORH counties)
Period for estimation of prescreening trend	1971-84	1970-8	1972-87	1971-85	1980-94
Selection method for last prescreening year	Opportunistic screening starts	Opportunistic screening starts	Last year before screening	Last year before screening	Last year before screening
Period for estimation of postscreening trend	1993-9	1995-2005	1996-2002	1998-2006	2000-6
Breast cancer incidence in final year of observation (per 100 000 women)					
Screened age group:					
Observed (regression analysis)	278	318/375*	291	328	303
Expected (regression analysis)	197	236/236*	211	242	213
Observed/expected	1.41	1.35/1.59	1.38	1.35	1.42
Exceeded age for screening:					
Observed (regression analysis)	278	401/442*	317	303 (1998)	246
Expected (regression analysis)	277	498/522*	289	338 (1998)	289
Observed/expected	1.01	0.81/0.85	1.10	0.90	0.85
Compensatory drop	No	Yes	No	Yes	Yes
Overdiagnosis (%) with CIS	NA	44	NA	NA	NA
Estimated overdiagnosis (%), assuming 10% CIS	57		53	46	52

AORH=Akershus, Oslo, Rogaland, and Hordaland; NA=not available; CIS=carcinoma in situ.
*Without/with CIS.

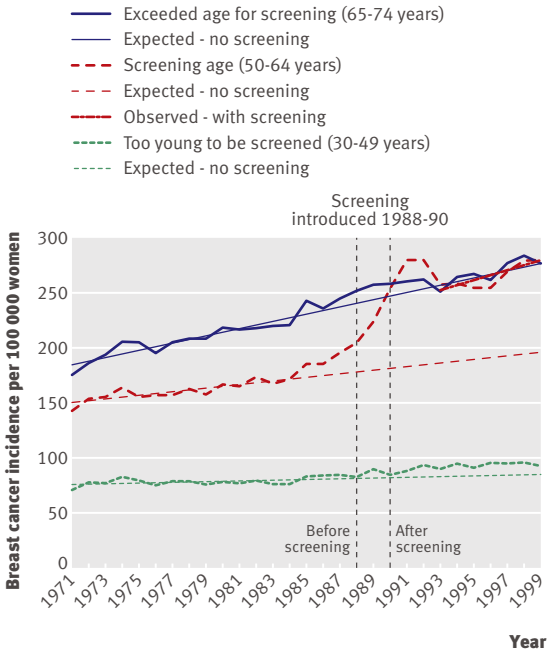

Fig 2 | Incidence of invasive breast cancer per 100 000 women in UK

Canada; New South Wales, Australia; Sweden; and parts of Norway (table).[12-16] (See web extra for data from an additional eight countries and reasons for exclusion.)

United Kingdom

Screening started in the UK in 1988 for women aged 50-64, with national coverage by 1990, and was expanded to women aged 65-70 in 2002.[17] Data from England and Wales covered 1971-99 in graphs with five year age groups.[12] These data were combined and the prescreening period defined as 1971-84, before opportunistic screening had influenced the background incidence (fig 2). The period 1993-9 was used to estimate the most recent trend. The increase in incidence of invasive cancer in women aged 50-64 was 41% above the expected rate, interpreted as overdiagnosis as there was no compensatory drop in the older age groups (fig 2). The incidence in younger age groups (30-49 years) increased by 7% over expected rates and in older age groups (65-74 years) by 1% over expected rates. No data were available for carcinoma in situ, but assuming that 10% of the diagnoses in a population offered screening are for carcinoma in situ,[10 11] overdiagnosis would be 57% (table).

More recent data (1995-2003) have been published,[17] but only for screened age groups. Incidence continues to increase.

Manitoba, Canada

No national data were found for Canada. In Manitoba, elective screening has been available since the late 1970s, with formal implementation in 1995 for women aged 50-69.[13] A study compared incidence up to 1999.[13] More recent data were received from the author (fig 3). As the incidence of carcinoma in situ started to increase in 1979, corresponding to the availability of elective screening, the prescreening period was defined as 1970-8. The period 1995-2005 was used to estimate the trend after screening. In the invited age group the incidence for invasive cancer was 35% above the expected rate, and when carcinoma in situ was included it was 59% higher. The total rate for the age group 70-84 was 15% below expected, but for the age group 35-49 it was 32% below expected, which suggests that causes other than screening could have contributed to the drop among previously screened women.

In the last observation year the 59% increase (including carcinoma in situ) in women aged 50-69 corresponds to 140 extra breast cancer diagnoses per 100 000 women, and the 15% decline in women aged 70-84 corresponds to 80 fewer breast cancer diagnoses per 100 000 women. In Manitoba, 2.3 times as many women are aged 50-69 than are aged 70-84,[9] and 75% (=(140×2.3-80)/(140×2.3)) of the increase is therefore uncompensated. A conservative estimate of overdiagnosis is therefore 59%×75%=44%.

New South Wales, Australia

National data on prescreening rates were not presented for Australia.[18] The introduction of screening varied from state to state, and follow-up was short.

For New South Wales, where screening was introduced during 1988-95, a graph showed an increase of 55% for invasive cancer over expected rates in women aged 50-69.[14] When the prescreening period was defined as 1972-87 and the period 1996-2002 was used to estimate the trend after screening, this age group showed an increase of 38% over expected rates

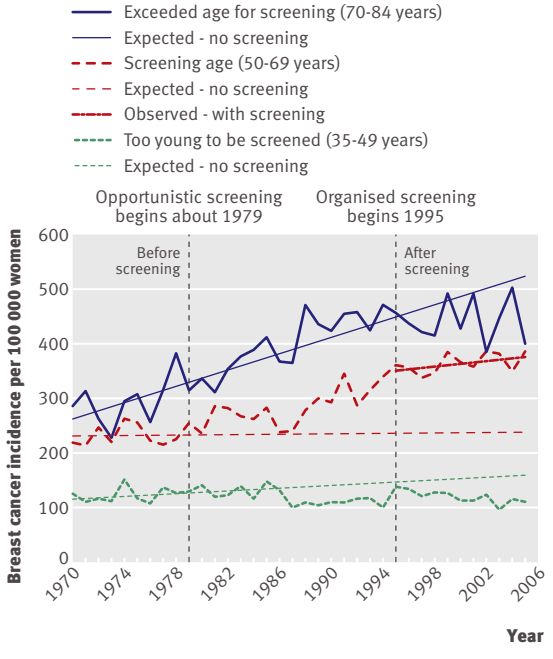

Fig 3 | Incidence of invasive breast cancer and carcinoma in situ per 100 000 women in Manitoba, Canada

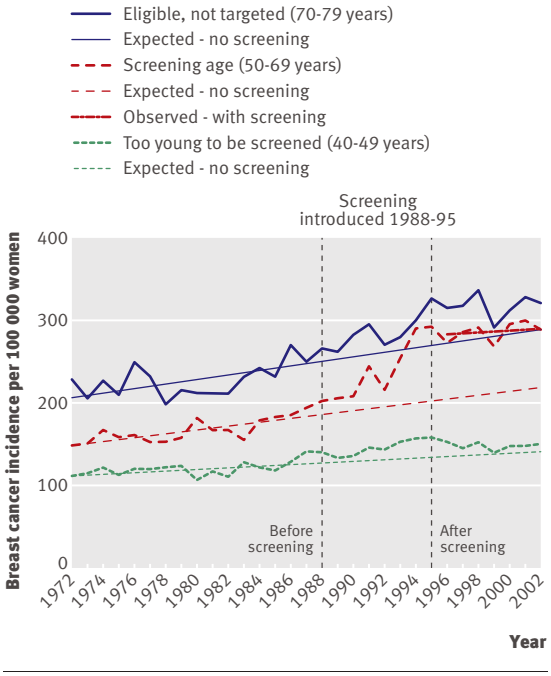

Fig 4 | Incidence of invasive breast cancer per 100 000 women in New South Wales, Australia

(fig 4). Among women too young to be screened the increase in incidence was constant (fig 4). Women aged more than 70 were eligible but not targeted. No compensatory drop was observed; the incidence was in fact larger than expected. Overdiagnosis including carcinoma in situ was therefore estimated at 53% (table).

A similar development was seen in South Australia, but the prescreening period was indicated as one data point, which precluded estimation of prescreening trends.[19]

Sweden

Women in a few areas of Sweden participated in screening trials from 1969; nationwide screening started in 1986, and in 1998 almost all eligible women had been offered screening.[20] For various counties in 1999, eight different targeted age ranges were described[20]; the broadest was 40-74 years and the most common was 50-69 years. A study reported an increase in invasive cancer after screening of 69% above expected rates in women aged 50-59 and 27% in women aged 60-69.[15] After adjustment for lead time, with estimates varying from 1.6 to 3.0 years, the increases in 2000 were 54% and 21%, respectively.[15] Another report[21] showed similar increases, without a compensatory drop in older age groups, whereas a third report noted a drop in incidence of 12% in those aged more than 75, and no change for women aged 70-74.[22]

Data up to 2006 were received from one of the authors (fig 5).[22] The meta-analysis focused on the age group 50-69, as this is the only group offered screening in all regions. Using the prescreening period as 1971-85 and the period 1998-2006 to estimate the trend after screening, the estimated increase for invasive cancer over expected rates was 35%, or 86 additional breast cancers per 100 000 women in the last observation year. A constant increase in incidence was seen among women too young to be screened (fig 5). A drop occurred among women aged 70-84, but incidence approached the expected rate at the end of the observation period (fig 5). In the middle of the interval after screening had started in 1998, 10% fewer invasive breast cancers were detected than expected, or 35 fewer cancers per 100 000 women. Eighty eight per cent of the increase was therefore uncompensated. Despite using data when the compensatory decline was largest (rather than from the last observation year), this adjustment only changed the estimate of overdiagnosis for invasive breast cancer from 35% to 31%. When carcinoma in situ was included overdiagnosis was 46% (table).

Norway

Screening was introduced in Norway in 1995-6 for women aged 50-69, but only in 40% of the population (Akershus, Oslo, Rogaland, and Hordaland counties; fig 6), and in the rest of Norway from 1999, gaining national coverage in 2004 (fig 7).[16] Attendance was good (75-77%).[16 22] As screening was fully implemented in the other counties in 2004, overdiagnosis was not estimated for these areas, although the data are presented graphically for comparison (fig 7). In Akershus, Oslo, Rogaland, and Hordaland, a peak in prevalence for invasive breast cancer was followed by stable levels, above prescreening rates in the screened age group.[16 22] Screening is generally offered to women aged 50-69, but about 50% of those aged 70-74 were probably screened,[23] and incidence initially increased by 30% in this age group and then decreased to

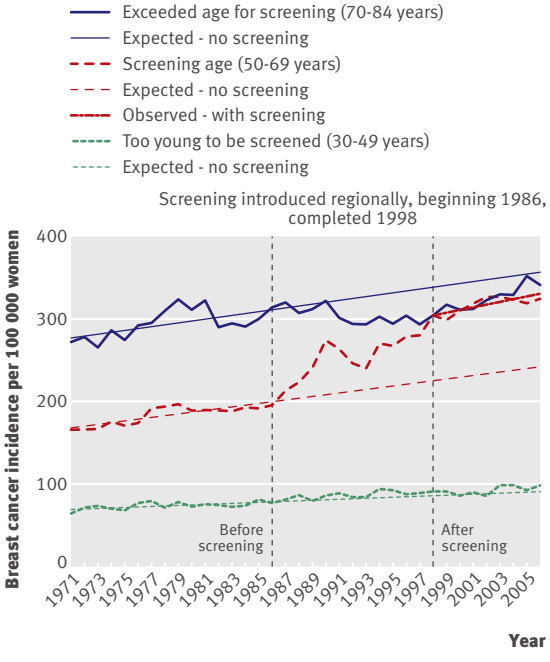

Fig 5 | Incidence of invasive breast cancer per 100 000 women in Sweden

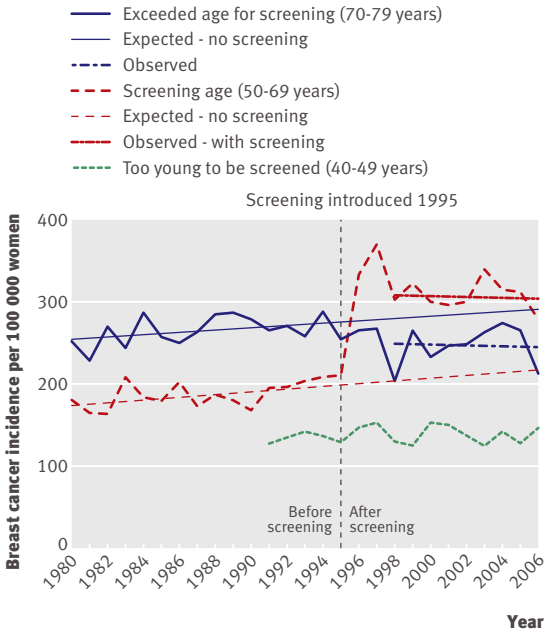

Fig 6 | Incidence of invasive breast cancer per 100 000 women in Akershus, Oslo, Rogaland, and Hordaland counties in Norway

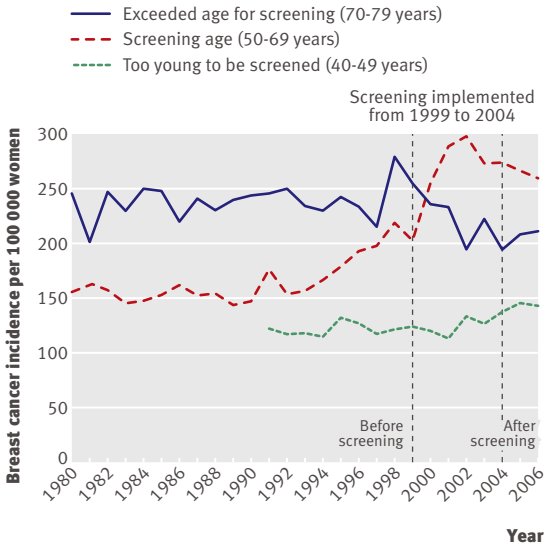

Fig 7 | Incidence of invasive breast cancer per 100 000 women in counties other than Akershus, Oslo, Rogaland, and Hordaland in Norway

prescreening levels. The incidence in women aged 20-50 and more than 74 was stable. Another study reported similar increases but had shorter follow-up.[22]

Additional data were received from one of the authors.[22] The age group 50-69 years was considered as screened. The prescreening period was defined as 1980-94 and the period 2000-6 was used to estimate the trend after screening. The increase in invasive breast cancer was estimated as 42% above expected rates, or 90 additional breast cancers per 100 000 women in the last observation year. Among women too young to be screened the increase in incidence was constant, but data for this group were only available divided into counties from 1991 (fig 6). A 15% drop was seen among women aged 70-79, but a similar drop was also observed in the rest of Norway before screening was fully implemented (fig 7). The drop was conservatively considered as compensatory. The 15% fewer invasive breast cancers correspond to 43 fewer cancers per 100 000 women. This means that 86% of the increase was uncompensated for, or that overdiagnosis was 37%. When carcinoma in situ was included overdiagnosis was 52% (table).

Meta-analysis

The total overdiagnosis of breast cancer in publicly available mammography screening programmes (including carcinoma in situ) was 52% (95% confidence interval 46% to 58%; fig 8). Heterogeneity was moderate (I^2=59%).

DISCUSSION

In populations offered organised screening for breast cancer, overdiagnosis (the detection of cancers that do not cause death or symptoms) was 52%. Carcinoma in situ was included in this estimate, as it is generally treated in the same way as invasive breast cancer[1 2 24]; the overdiagnosis for invasive breast cancer only was 35% (95% confidence interval 29% to 42%).

We took account of the increasing background incidence by comparing the observed rates of breast cancer with the expected rates for the last year of observation, using projected incidence rates from prescreening trends. Our assumption of a constant, linear increase in the background incidence was supported by data from age groups that were too young to be screened, as agreement between projected and observed rates was good (figs 2-5). Another indication that our assumption was reasonable is that the incidence of breast cancer only deviated from a linear increase around the time of the introduction of screening. This was the case in all included areas, even though screening was introduced at different times (from 1979 in Manitoba to 1995 in Norway). It is therefore unlikely that changes in risk factors or cohort effects could explain the non-linear increases in incidence of breast cancer that occurred with the introduction of screening.

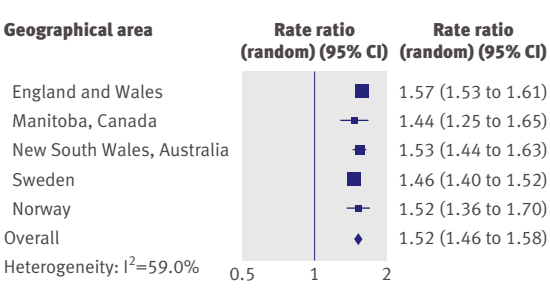

Fig 8 | Meta-analysis of overdiagnosis of breast cancer (including carcinoma in situ) in publicly available mammography screening programmes

WHAT IS ALREADY KNOWN ON THIS TOPIC

Screening for cancer detects inconsequential cancers and leads to overdiagnosis and overtreatment

A Cochrane review of the randomised trials of mammography screening documented 30% overdiagnosis

Overdiagnosis in publicly organised mammography screening programmes has not been evaluated systematically

WHAT THIS STUDY ADDS

Overdiagnosis of breast cancer in a population offered organised mammography screening was 52%

This extent of overdiagnosis equates to one in three breast cancers being overdiagnosed

Manitoba had substantial opportunistic screening before organised screening was introduced,[13] but we avoided this bias by estimating the prescreening trends from periods when there were few diagnoses of carcinoma in situ.

The trend after implementation of screening was estimated under the assumption that screening leads to a higher incidence level that increases at about the same rate as the background incidence did before screening.[25] Our data support this assumption (figs 2-6).

As we have data on long follow-up it is unlikely that the increasing incidence in the screened age group will be compensated for later on. Screening theory implies that a compensatory drop would be apparent shortly after women leave the screening programme and thus after comparatively short follow-up.[25]

Not all women in all areas passed from the screened age group to the previously screened age group within our observation period. In England and Wales, however, practically all women aged 65-74 would have been offered screening previously at the end of our observation period, but we did not find a compensatory drop in incidence of breast cancer (fig 2).

Some authors use statistical models to adjust their estimate of overdiagnosis for lead time (increased incidence because of advancement of the time of diagnosis).[26-30] This approach is problematic as all models carry a high risk of bias[31] because they are based on unverified assumptions, and as the choice of variables is crucial—for example, high estimates of lead time result in low estimates of overdiagnosis.[31] Estimates of lead time varied between 1.6 and 4 years and even differed in articles by the same authors.[15 26 27 29 30]

The recent decline in the use of hormone replacement therapy after evidence that it causes breast cancer[32] is a possible explanation for the reduction in incidence observed in the United States from 2002, in particular as such a decline did not occur in women below 50 years of age.[33] We did not, however, see similar declines in the countries we examined, and the declining use of mammography screening in the United States has also been suggested as an explanation.[34]

In Norway the effect of screening was separated from that of hormone replacement therapy use, as incidence trends in regions with and without screening could be compared at the same calendar times. Although use of hormone replacement therapy is likely to be similar, a noticeable increase occurred in invasive cancer with the introduction of screening, both in the Akershus, Oslo, Rogaland, and Hordaland counties and in the remaining counties of Norway (figs 6 and 7), and in the other regions we examined (figs 2-4).

Conclusion

We estimated 52% overdiagnosis of breast cancer in a population offered organised mammography screening—that is, one in three breast cancers is overdiagnosed.

We thank Alain Demers (Department of Epidemiology and Cancer Registry, CancerCare Manitoba) for supplying data for Manitoba, Per-Henrik Zahl (Folkehelseinstituttet, Oslo) for supplying data for Sweden and Norway, and Gilbert Welch for helpful comments on the manuscript.
Contributors: PCG and KJJ conceived the study, developed the methods, extracted data and carried out the analyses. KJJ did the searches, contacted authors, and wrote the first draft of the manuscript, which was revised by PCG. Both authors are guarantors.
Funding: None.
Competing interests: None declared.
Ethical approval: Not required.

1 Vainio H, Bianchini F. *IARC handbooks of cancer prevention vol 7: breast cancer screening*. Lyon: IARC Press, 2002.
2 Welch GH. *Should I be tested for cancer?* California: University of California Press, 2004.
3 Nielsen M, Thomsen JL, Primdahl S, Dyreborg U, Andersen JA. Breast cancer atypia among young and middle-aged women: a study of 110 medicolegal autopsies. *Br J Cancer* 1987;56:814-9.
4 Welch HG, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast. *Ann Intern Med* 1997;127:1023-8.
5 Gøtzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;355:129-34.
6 Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2006;(4):CD001877.
7 Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ* 1988;297:943-8.
8 Jørgensen KJ, Klahn A, Gøtzsche PC. Are benefits and harms in mammography screening given equal attention in scientific articles? A cross-sectional study. *BMC Med* 2007;5:12.
9 Statistics Canada. 2008. www12.statcan.ca/english/census06/data/topics/RetrieveProductTable.cfm?Temporal=2006&PID=88984&GID=838031&METH=1&APATH=3&P-=2006&PID=88984&GID=838031&METH=1&APATH=3&PTY-=88984&GID=838031&METH=1&APATH=3&PTYPE=88971&THE-=838031&METH=1&APATH=3&PTYPE=88971&THEME=66&AID=&-=1&APATH=3&PTYPE=88971&THEME=66&AID=&FREE=0&FO-=3&PTYPE=88971&THEME=66&AID=&FREE=0&FOCUS=&-=88971&THEME=66&AID=&FREE=0&FOCUS=&VID=0&GC=99&G-=66&AID=&FREE=0&FOCUS=&VID=0&GC=99&GK=NA&RL=0&TPL=-=0&FOCUS=&VID=0&GC=99&GK=NA&RL=0&TPL=RETR&-=&VID=0&GC=99&GK=NA&RL=0&TPL=RETR&SUB=0&d1=0.
10 Lindgren A, Holmberg L, Thurfjell E. The influence of mammography screening on the pathological panorama of breast cancer. *APMIS* 1997;105:62-70.
11 NHS. *Screening for breast cancer in England: past and future*. Publication No 61. Department of Health: NHS cancer screening programmes, 2006. www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp61.html.
12 Johnson A, Shekhdar J. Breast cancer incidence: what do the figures mean? *J Eval Clin Pract* 2005;11:27-31.
13 Demers AA, Turner D, Mo D, Kliewer EV. Breast cancer trends in Manitoba: 40 years of follow-up. *Chronic Dis Can* 2005;26:13-9.
14 Chiu C, Morrell S, Page A, Rickard M, Brassil A, Taylor R. Population-based mammography screening and breast cancer incidence in New South Wales, Australia. *Cancer Causes Control* 2006;17:153-60.

15 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.
16 Hofvind S, Sorum R, Haldorsen T, Langmark F. [Incidence of breast cancer before and after implementation of mammography screening]. *Tidsskr Nor Laegeforen* 2006;126:2935-8.
17 NHS Health and Social Care Information Centre. Breast screening programme, England: 2004-05. www.cancerscreening.nhs.uk/breastscreen/breast-statistics-bulletin-2004-05.pdf.
18 Giles G, Amos A. Evaluation of the organised mammographic screening programme in Australia. *Ann Oncol* 2003;14:1209-11.
19 Luke C, Priest K, Roder D. Changes in incidence of in situ and invasive breast cancer by histology type following mammography screening. *Asian Pac J Cancer Prev* 2006;7:69-74.
20 Sjönell G, Ståhle L. [Health checks with mammography do not reduce breast cancer mortality]. *Läkartidningen* 1999;96:904-13.
21 Hemminki K, Rawal R, Bermejo JL. Mammographic screening is dramatically changing age-incidence data for breast cancer. *J Clin Oncol* 2004;22:4652-3.
22 Zahl PH, Strand BH, Mæhlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004;328:921-4.
23 Møller B, Weedon-Fekjær H, Hakulinen T, Tryggvadottir L, Storm HH, Talback M, et al. The influence of mammographic screening on national trends in breast cancer incidence. *Eur J Cancer Prev* 2005;14:117-28.
24 Douek M, Baum M. Mass breast screening: is there a hidden cost? *Br J Surg* 2003;90:44-5.
25 Boer R, Warmerdam P, de Koning H, van Oortmarssen G. Extra incidence caused by mammographic screening. *Lancet* 1994;343:979.
26 Svendsen AL, Olsen AH, von Euler-Chelpin M, Lynge E. Breast cancer incidence after the introduction of mammography screening: what should be expected? *Cancer* 2006;106:1883-90.
27 Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F, 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.
28 Olsen AH, Jensen A, Njor SH, Villadsen E, Schwartz W, Vejborg I, et al. Breast cancer incidence after the start of mammography screening in Denmark. *Br J Cancer* 2003;88:362-5.
29 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.
30 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.
31 Zahl PH, Jørgensen KJ, Mæhlen J, Gøtzsche PC. Biases in published estimates of overdetection due to screening with mammography. *Lancet Oncol* 2008;9:199-201.
32 Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
33 Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, et al. The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* 2007;356:1670-4.
34 Thorat MA. What caused the decline in US breast cancer incidence? *Nat Clin Pract Oncol* 2008;5:314-5.

Accepted: 30 January 2009

BMC Women's Health

Research article

Open Access

Overdiagnosis in organised mammography screening in Denmark. A comparative study

Karsten J Jørgensen*[1], Per-Henrik Zahl[2] and Peter C Gøtzsche[1]

Address: [1]The Nordic Cochrane Centre, Rigshospitalet, Dept 3343, Blegdamsvej 9, DK-2100 Copenhagen, Denmark and [2]Norwegian Institute of Public Health, PO Box 4404, Nydalen, N-0405, Oslo, Norway

Email: Karsten J Jørgensen* - kj@cochrane.dk; Per-Henrik Zahl - per-henrik.zahl@fhi.no; Peter C Gøtzsche - pcg@cochrane.dk

* Corresponding author

Published: 22 December 2009

BMC Women's Health 2009, **9**:36 doi:10.1186/1472-6874-9-36

Received: 18 September 2009
Accepted: 22 December 2009

This article is available from: http://www.biomedcentral.com/1472-6874/9/36

Abstract

Background: Overdiagnosis in cancer screening is the detection of cancer lesions that would otherwise not have been detected. It is arguably the most important harm. We quantified overdiagnosis in the Danish mammography screening programme, which is uniquely suited for this purpose, as only 20% of the Danish population has been offered organised mammography screening over a long time-period.

Methods: We collected incidence rates of carcinoma in situ and invasive breast cancer in areas with and without screening over 13 years with screening (1991-2003), and 20 years before its introduction (1971-1990). We explored the incidence increase comparing unadjusted incidence rates and used Poisson regression analysis to compensate for the background incidence trend, variation in age distribution and geographical variation in incidence.

Results: For the screened age group, 50 to 69 years, we found an overdiagnosis of 35% when we compared unadjusted incidence rates for the screened and non-screened areas, but after compensating for a small decline in incidence in older, previously screened women. Our adjusted Poisson regression analysis indicated a relative risk of 1.40 (95% CI: 1.35-1.45) for the whole screening period, and a potential compensatory drop in older women of 0.90 (95% CI: 0.88-0.96), yielding an overdiagnosis of 33%, which we consider the most reliable estimate. The drop in previously screened women was only present in one of the two screened regions and was small in absolute numbers.

Discussion: One in four breast cancers diagnosed in the screened age group in the Danish screening programme is overdiagnosed. Our estimate for Denmark is lower than that for comparable countries, likely because of lower uptake, lower recall rates and lower detection rates of carcinoma in situ.

Background

Overdiagnosis in cancer screening is defined as the detection of cancers that would otherwise not have been detected in the remaining life-span of the individuals [1]. It is mainly caused by the detection of slow-growing cancers that do not manifest clinically before people die from

other causes [2], but may also be due to identification of borderline malignancies, or cancers that were bound to regress [3,4].

Overdiagnosis is arguably the most important harm of screening, as healthy people are being diagnosed with and treated for cancer unnecessarily, which carries great personal costs, both physically and psychologically [2].

Overdiagnosis is an unavoidable consequence of mammography screening [2]. It is also well documented for other cancers, e.g. lung cancer, neuroblastoma and prostate cancer [5]. Although about 60% of men in their 60's have cancer lesions in their prostate in autopsy studies, the observed incidence is much smaller and the lifetime risk of dying from prostate cancer is only 3% [5]. Some of these otherwise undetected cancers will become diagnosed if screening is implemented, and this is an important reason that screening for prostate cancer is discouraged in many Western countries [5]. Overdiagnosis is also a problem with mammography screening, but it has been been omitted in most information material intended to help women make informed decisions about participation [6-8].

Overdiagnosis in mammography screening has been documented in systematic reviews of the randomised trials. Screening led to a 31% increase in the use of breast cancer surgery, which included a 20% increase in the use of mastectomies [9]. We have recently quantified the extent of overdiagnosis in breast cancer screening programmes in Manitoba, New South Wales, Norway, Sweden and the UK and found 52% overdiagnosis, including carcinoma in situ lesions (CIS) [10]. There were large and sustained increases in breast cancer incidence when screening was introduced, with only small or absent compensatory decreases among older, previously screened women. A compensatory decrease in incidence is required, if the incidence increase in the screened group is due only to advancement of the time of diagnosis (lead-time) [11]. Data on CIS lesions are often lacking in articles that describe indicence trends [10], although CIS is mainly diagnosed through screening and contributes substantially to overdiagnosis, as these lesions currently constitute 20% of breast cancers detected at screening in the UK and more in the USA [12,13], and although less than half of them progress to invasive breast cancer [2]. A limitation of our systematic review was that there were no concomitant control groups without screening and we therefore used linear projections of pre-screening rates to estimate the overdiagnosis [10].

Estimating overdiagnosis can be particularly difficult in countries that do not have an organised screening programme and therefore have no well-defined pre- and post-screening period, e.g. the USA. Also, in the USA, there is no control group without screening, as all areas have screening in private practices. Comparisons with younger and older age groups is also unreliable in this case, as women are recommended to be screened from they are very young, and as there is no upper age limit. Contrary to this, the Danish breast cancer screening programme provides a unique opportunity for estimating overdiagnosis, as there has been a period of 17 years (1991-2007) where only about 20% of potentially eligible women have been invited to screening with mammography, in two separate administrative regions, with a well-defined target age group and starting point. Partly because of an intensive debate about the balance between the benefits and harms of screening, the Danish administrative regions have prioritised their resources differently. A national screening policy has now been adopted and is currently being implemented.

We compared breast cancer incidence in areas with and without organised screening. We studied whether the increase in incidence in women who were offered screening was compensated by a drop in breast cancer incidence when the women passed the age limit for screening, and compared with the development in women in the same age group in areas without screening, and with that in younger women. We corrected the estimates of overdiagnosis for differences in age distribution, geographical differences in pre-screening incidence rates and changes in background incidence with Poisson regression analyses.

Methods

We retrieved data on breast cancer incidence in females during 1971-2003 from the Danish Cancer Registry at the National Board of Health (data were not yet available for 2004-2007). The data we received were aggregated by geographical region and into 5-year age groups. Data on population size for each year, region and age group were obtained from Statistics Denmark [14]. We included carcinoma in situ, as these lesions are treated surgically as if they were invasive cancers.

Organised mammography screening of women aged 50-69 years began April 1st 1991 in Copenhagen municipality, November 1st 1993 in Funen County, and June 1st 1994 in Frederiksberg municipality. The Frederiksberg programme, which comprised comparatively few women, was incorporated into the Copenhagen programme January 1st 1997 [15]. In 2003, there were 115,270 women aged 50-69 years old in the screened areas (54,933 in Copenhagen and Frederiksberg and 60,337 in Funen), and 551,778 in areas without organised screening [14].

The attendance rates were 63% in Copenhagen and 83% in Funen in the fourth screening round [16], which are

lower than in Sweden, Norway, Finland and the UK [2]. The recall rates per incidence round were 4.3% in Copenhagen and 1.3% in Funen [15], which are also lower than in comparable countries.

The population in Copenhagen and Funen is comparable with the rest of Denmark regarding age distribution and socioeconomic status. Copenhagen is the largest city, but the second largest is in the non-screened areas, and there are rural areas in Funen County, like in the rest of Denmark. The Danish population is one of the most homogenous in the world.

We first explored the incidence increase using the observed rates, without adjustments. We calculated the average incidence rate ratio between screened and non-screened areas for the screened age group. We then subtracted the compensatory decline in older, previously screened women (ages 70-79 years) to estimate overdiagnosis, after calculating the absolute number of diagnoses that the incidence rates corresponded to. Two years after screening was introduced, women aged 70 and 71 would have been offered screening once, as women are called for screening every second year. After four years, women aged 72 and 73 would have been offered screening once, and women aged 70 and 71 would have been offered screening twice. After 10 years, all women would have been offered screened previously (which is in 2001 in Copenhagen and in 2003 in Funen), with the younger ages having been offered screening more than once. We would therefore expect to see a trend towards declining incidence in the age group 70-79 years, beginning early on after the onset of screening. We also used simple linear regression to estimate incidence trends before and after the introduction of screening [10].

We used Poisson regression analyses to obtain more reliable estimates of overdiagnosis, with confidence intervals, and adjusted for differences in age distribution using 5-year age groups, and for the fact that Funen introduced screening 3 years after Copenhagen, and for geographical differences in incidence, using the pre-screening period as reference.

We also used Poisson regression analysis to quantify the compensatory drop in incidence in women ages 70-79 years. The Poisson regression model was adjusted only for age and included a trend parameter, starting in 1998, because 7 in 10 women in Copenhagen in this age group, and 4 in 10 women in Funen, would then have been offered screening previously. We used the incidence in the rest of Denmark as reference to compensate for the increasing background incidence in this age group.

The statistical analyses were performed using Egret version 2.0.3 and graphs were made in Microsoft Excel 2000.

Results

Data were available from 13 years of organised mammography screening (1991-2003). In women aged 50-69 years, 5,189 cases of breast cancer (of which 6% were CIS) were diagnosed during 1,342,836 woman-years in areas offering screening (average 386/100,000 woman-years), and 17,686 cases (3% CIS) were diagnosed during 6,191,609 woman-years in areas not offering screening (average 286/100,000 woman-years) (Table 1).

Among women in the screened age group, pre-screening breast cancer incidence increased at a stable rate from 1971-1990 (Fig. 1), with slightly higher rates in the screened areas than in the non-screened areas (average 214 vs. 198 breast cancers per 100,000 person-years) (Table 1). There were also slightly higher rates in the screened areas in the age groups 35-49 years and 70-79 years in this period (Table 1).

The breast cancer incidence doubled in Copenhagen in 1991 and in Funen in 1994, when screening was introduced (Fig. 1). After the first round of screening, the breast cancer incidence in screened areas was about 30% higher than in the non-screened areas, and more when compared with the expected incidence projected from the pre-screening incidence. Using a linear regression analysis for the screened areas for the incidence rounds 1996-2003, there were 36% more breast cancers than expected in 2003 in the screened areas (Fig. 1). In the non-screened areas, the breast cancer incidence was also somewhat higher than expected (Fig. 1).

The screening activity in the age group 50-69 years is reflected in the detection rates of CIS (Fig. 2). There were increases of several hundred per cent in the screened areas, both compared to the rates in the non-screened areas, and to expected rates (Fig. 2). Rates of CIS increased somewhat in the screened areas before the introduction of organised screening, beginning in 1988 (Fig. 2), likely because of opportunistic screening.

In the age group 70-79 years, breast cancer incidence also increased steadily from 1971-1990 (Fig. 3). Breast cancer incidence in Copenhagen increased more than expected after the introduction of screening in 1991 (Fig. 3). In Funen, breast cancer incidence also increased after the introduction of screening, but there was a drop towards the end of the observation period (Fig. 3). Breast cancer incidence in the non-screened areas also increased more than expected after 1991 (Fig. 3).

In women aged 35-49 years, breast cancer incidence increased steadily in the period 1971-1990. It then stabilized in Copenhagen and the non-screened areas, but not in Funen (Fig. 4).

Table 1: Number of breast cancers, number of women, and incidence rates in screened and non-screened areas, before and after screening started, and during the last three years of observation.

		Screened areas			Non-screened areas		
		1971-1990 No screening	1991-2003 Screening	2001-3 Screening	1971-1990	1991-2003	2001-3
Breast cancers	35-49 years	2,110	1,684 (4% CIS)	381 (3% CIS)	8,668	7,228 (5% CIS)	1,741 (4% CIS)
	50-69 years	5,846	5,189 (6% CIS)	1,282 (6% CIS)	16,263	17,686 (3% CIS)	4,922 (2% CIS)
	70-79 years	3,258	2,058(3% CIS)	383 (2% CIS)	7,256	6,483 (2% CIS)	1,767 (2% CIS)
Person years	35-49 years	1,759,614	1,317,024	314,322	7,827,731	6,038,527	1,412,069
	50-69 years	2,737,925	1,342,836	326,946	8,223,810	6,191,609	1,565,967
	70-79 years	1,195,296	606,034	117,077	2,743,410	2,100,884	481,608
Per 100,000	35-49 years	120	128	121	111	120	123
	50-69 years	214	386	392	198	286	314
	70-79 years	273	340	327	264	309	367

In the 13 years of screening, there were 1,343 extra breast cancers detected in the screened areas compared with the non-screened areas in women aged 50-69 years, when comparing the unadjusted incidence rates (386 vs. 286 breast cancers per 100,000 women in 1,342,836 woman years, Table 1), and there were 182 *extra* cancers detected in the screened areas collectively, compared with the non-screened areas, among women aged 70-79 years, i.e. no compensatory drop (Table 1). The 1,343 extra cancers are equivalent to 35% overdiagnosis in the screened population (Table 2). If we compare only the years 2001-3, where there was a drop in incidence in Funen among women aged 70-79 years, there were 255 extra cancers detected in the age group 50-69 years, 47 of which may have been compensated. This conservative analysis means that there were 208 uncompensated extra cancers in the screened population, or 19% overdiagnosis (Tables 1 and 2). However, the data set is small and random fluctuation might therefore explain that this estimate is lower than that for the main analysis.

Poisson regression analyses

In our Poisson regression analysis, where we took potential biases into account, we found a risk ratio of 2.07 (95% CI: 1.94-2.21) for the first rounds, and 1.34 (95% CI: 1.29-1.40) in the following period (Table 3). The weighted average was 1.40 (95% CI: 1.35-1.45).

We found a risk ratio of 0.90 (95% CI: 0.85-0.96) in women aged 70-79 years in Copenhagen and Funen combined in the period 1998-2003, using Poisson regression analysis. Using the same model for the age group 70-74 years, where a compensatory drop would appear first and be most pronounced, we found a risk ratio of 0.89 (95% CI: 0.80-0.96) (Table 3). We used Poisson regression to test for a trend towards a linearly accellerating drop in incidence in women 70-79 years, as such a trend would increase the likelihood that screening caused the decline. Again, we chose 1998 as our starting point and repeated the analysis for the age group 70-74 years to increase the chance of detection. We did not find such a trend, neither for the age group 70-79 years (P = 0.50), nor for the age group 70-74 years (P = 1.00), the annual percentage change was 0.37% (95% CI: 0.30-0.45).

Forty percent more breast cancers in the screened population and a 10% compensatory drop in previously screened women means that there were 5, 189 - (5,189/1.40) = 1,483 extra breast cancers detected in the period 1991-2003, and that 2,058 × 0.1 = 206 of these were later compensated (Table 1), which gives 33% overdiagnosis.

Discussion

When we adjusted for regional differences in incidence and age distribution, and compensated for a decline in incidence in older, previously screened women, we found 33% overdiagnosis in Denmark. This is lower than the 52% we estimated for other countries in our systematic review of organised screening programmes, which did not include Denmark [10]. The likely reasons for this are that

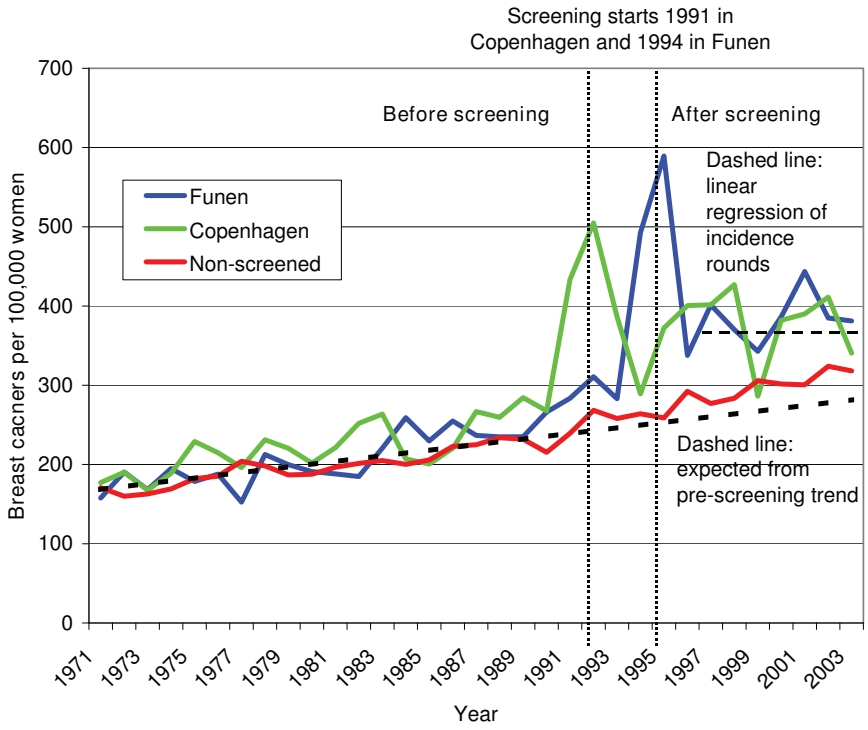

Figure 1
Unadjusted incidence of in situ and invasive breast cancers per 100,000 women ages 50-69 years in areas without mammography screening and in Copenhagen and in Funen.

the Danish screening programme has low recall rates, e.g. only 1.3% per round in Funen [15], low detection rates for CIS (6% of diagnoses, Table 1), because of a deliberately conservative attitude towards detection of microcalcifications [17], and comparatively low uptake, e.g. 63% in Copenhagen [16].

We did not take opportunistic screening in the non-screened areas into account in our Poisson regression analyses, and we might therefore have underestimated overdiagnosis. We abstained from this adjustment, as the level of opportunistic screening in the non-screened areas is difficult to estimate. Increasing rates of CIS indicate that there was opportunistic screening in Copenhagen and Funen from about 1988, but that there was little opportunistic screening in the rest of Denmark (Fig. 2). Another study confirmed our findings and estimated that opportunistic screening outside the organised programme covered only 10% of the women [18], and noted that it was difficult to differentiate between diagnostic and screening mammograms using available statistics. CIS has been reported to be poorly registered outside the screening programmes [17] and increased little (Figure 2), whereas the incidence of breast cancer overall increased more than expected in the non-screened areas from 1991 (Fig. 1). This increase is partly due to overdiagnosis caused by opportunistic screening, but could also partly reflect an increase in background incidence, although the stable incidence rates in the age group 35-49 years speak against this (Figure 4). As we compared screened and non-screened regions, general changes in the background incidence would not affect our estimate of overdiagnosis, which is a strength of our study, compared to using projections of pre-screening incidence rates in the screened regions only [10,17].

In 2003, at the end of our observation period, practically all women aged 70-79 years in Copenhagen and Funen had been offered screening several times at an earlier age and had therefore contributed to the observed incidence increase in the age group 50-69 years earlier on. In other words, the cohort of women aged 70-79 years in 2003 consisted entirely of women who were 60-69 years earlier on in our observation period and who had therefore been offered screening previously. In the absence of overdiagnosis, the incidence in the age group 70-79 years should therefore drop more and more with increasing follow-up. However, there was no drop in incidence rates in Copenhagen and it is contrary to screening theory that we failed to detect a trend towards progressively lower incidence

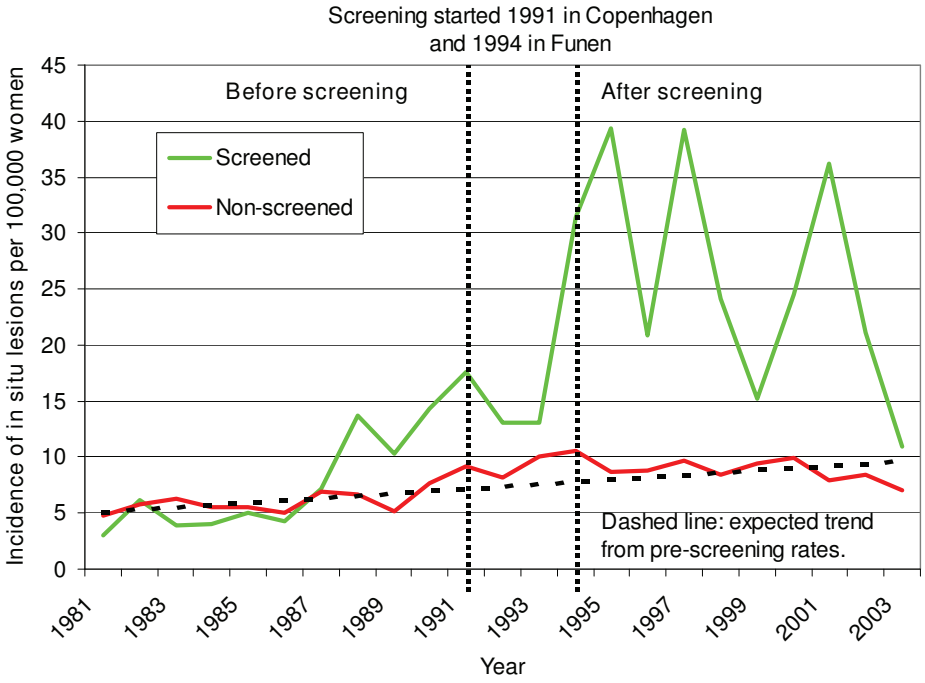

Figure 2
Unadjusted incidence of *in situ* lesions only, per 100,000 women ages 50-69 years, in areas without mammography screening and in Copenhagen and in Funen.

rates in previously screened women [11]. There was a drop in Funen after about 7 years with screening (Fig. 3), but as there were fewer breast cancers in Funen, and as the incidence increased more than expected in this age group in Copenhagen, the combined drop is negligible. A compensatory drop should have occurred earlier in Copenhagen than in Funen, which speaks against that the drop in Funen is related to screening. We consider it unlikely that longer follow-up would change these findings, as we were also unable to demonstrate important compensatory declines in countries with longer follow up than Denmark [10]. The data material is small enough that the drop in Funen could be a random fluctuation and more follow up is required to establish this. Until this is available, we therefore consider our conserative estimate of 19% overdiagnosis based on only the last observation years as less reliable.

In the screened areas, breast cancer incidence in the age group 50-69 years was higher than in the age group 70-79 years throughout the period with organised screening (compare Fig. 1 and 3). Mammography screening has dramatically changed the shape of the age-specific breast cancer incidence curve: it was increasing with age prior to screening, but now has a maximum in the age group 50-69 years. Furthermore, because there are many more women in the age group 50-69 years than in the age group 70-79 years (ratio 2.3 to 1, see Table 1), the drop in the incidence rate in the age group 70-79 must be much larger than the increase in the incidence rate in the age group 50-69 years, if all the extra breast cancers detected through screening are to be compensated. Even if no breast cancers were diagnosed in the age group 70-74 years in 2003, it would not account for the extra breast cancers detected in the screened age group. This is not compatible with common expectations of an average lead-time of 2-3 years [10,19,20] and indicates that a large part of the observed incidence increase must be due to other causes.

There were minor differences between the screened and non-screened regions in pre-screening incidence (Table 1 and 3). Such differences were expected, as large cities comprised a greater proportion of the screened areas, and as breast cancer incidence is generally higher in cities. There was also a higher incidence in the non-screened areas than expected from the linear projection of the pre-screening trend (Figure 2). This is partly due to opportunistic screening, as discussed above, but could also be due to other factors that increase the breast cancer risk, such as HRT (hormone replacement therapy). However, while it is likely that some women in the non-screened areas would seek opportunistic screening once it was available to oth-

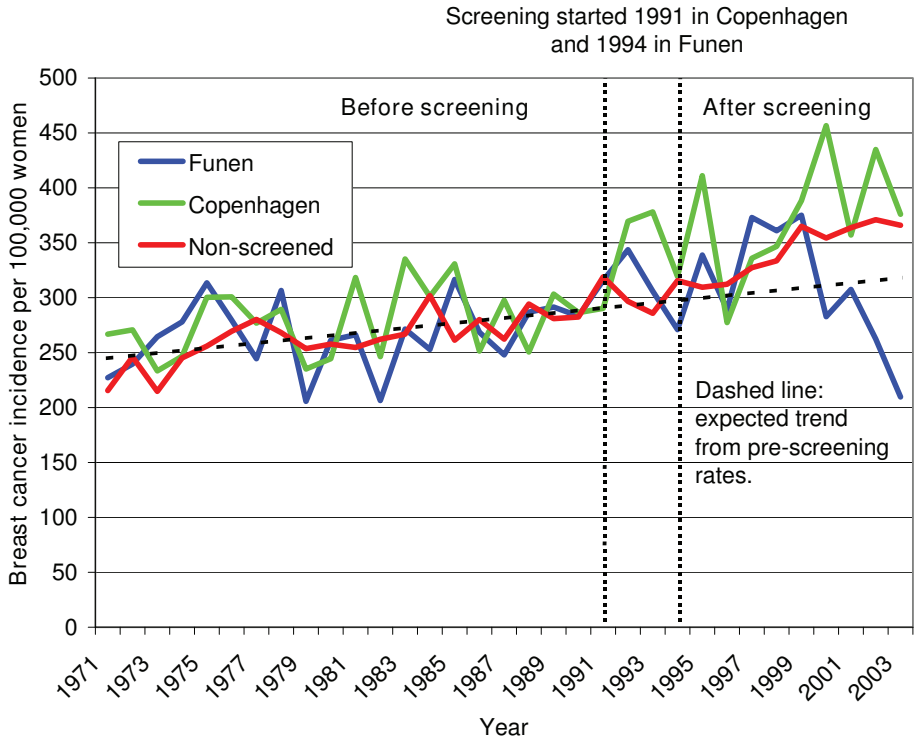

Figure 3
Unadjusted incidence of *in situ* lesions and invasive breast cancers per 100,000 women ages 70-79 years in areas without mammography screening and in Copenhagen and in Funen.

ers, there is no good reason why the increase should occur simultaneously with the introduction of screening if it were due to HRT. The findings of our recent systematic review on overdiagnosis also speak strongly against a general, non-screening related rise in the background incidence, over the pre-screening trend [10]. In the countries we studied, the abrupt increase in incidence always occurred simultaneously with screening, despite the fact that screening was introduced a decade apart in the various countries [10]. Further, the increase was by far most predominant in the invited age range, despite the fact that this also varied between countries [10].

We did not compare closed cohorts. Influx of patients to regions with screening could boost incidence in the screened age groups and in older age groups, leading to overestimation of overdiagnosis. However, the mobility between Danish regions in the age group 50-79 years is limited, and our findings are in good agreement with those from comparable countries such as Sweden and the UK, which has nation-wide screening and can be considered more similar to a closed cohort [19], and where we found an overdiagnosis of about 50% [10]. It could be argued that in the beginning of the observation period following the introduction of screening, we compare a screened cohort with a women who have never been offered screening. However, this is not so much a limitation of our cohorts as a choice that allows us to document if there was a trend towards a decline in the incidence rate of breast cancer, as the proportion of women in the age group 70-79 years that had previously belonged to the screened age cohort of women (women aged 50-69 years) increased. There was no such trend, and at the end of our observation period, we compared practically only currently screened women with previously screened women. Using only the last observation year would be unreliable, as the data would be prone to random fluctuations given our comparatively small sample size.

The absence or the small magnitude of a compensatory drop in previously screened women we found here, and in our previous review [10], questions the central premise of breast cancer screening. It shows that very little of the surplus of cancers observed in the screened age group can be due to advancement of the time of diagnosis (lead-time) for lethal cancers where screening might be beneficial. Previous estimates of average lead-times of 2-3 years [19,20] must therefore be wrong.

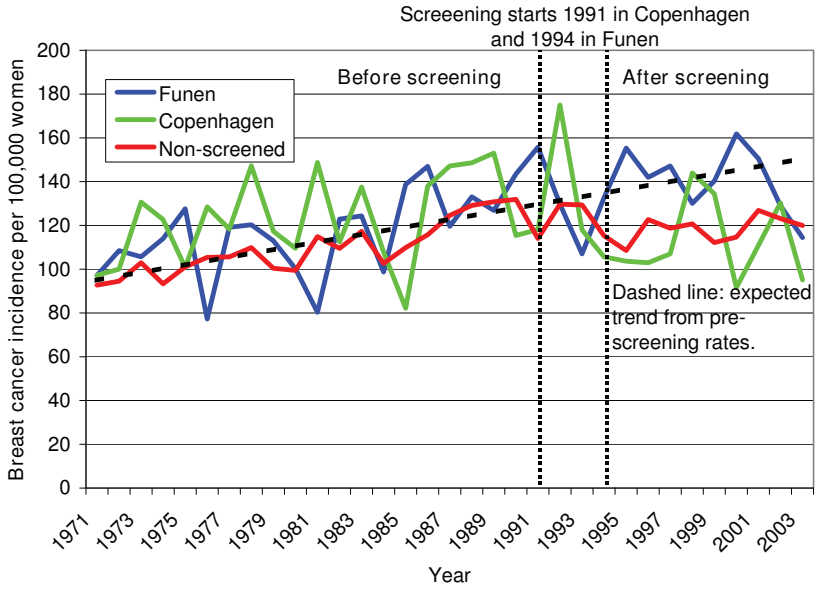

Figure 4
Unadjusted incidence of in situ and invasive breast cancers per 100,000 women ages 35-49 years in areas without mammography screening and in Copenhagen and in Funen.

It has been suggested by comparison of left- with right-sided irradiation that radiotherapy may double not only the mortality from heart disease, but also that from lung cancer [21], although technological improvements may have diminished these harms to some extent. Lung cancer is currently rivalling breast cancer as the leading cancer related cause of death among women in the Western world and heart disease is the major cause of death in women. It is therefore of interest that an effect of screening on all-cause mortality has not been demonstrated. The randomised trials were not powered to detect a difference in all-cause mortality, but as more than half a million women participated in the trials, this at least indicates a limited absolute benefit of the intervention [9].

More surprisingly, there is no indication that screening lowers all-cancer mortality, including breast cancer mortality. The relative risk was 1.00 (95% CI 0.96-1.05) in the randomised trials [9], although with the commonly stated 30% reduction in breast cancer mortality with screening, the expected relative risk for all-cancer mortality would be 0.95, which is below the confidence interval of what was actually found [9].

Svendsen et al. have previously concluded that the Danish data do not provide evidence of overdiagnosis of invasive breast cancer, or that it was of limited magnitude [17]. The authors reported that the observed incidence rate in the screening period in Copenhagen and Funen, considered separately from each other, was within the 95% confidence interval of the expected rate, which they projected from the pre-screening rates using regression analysis. Our linear regression analysis of the pre-screening rates in the two screening regions combined indicated that a total of

Table 2: Number of extra cancers in screened women (ages 50-69 years), number of extra cancers compensated in previously screened women (ages 70-79 years), and unadjusted estimates of overdiagnosis in Denmark.

	1991-2003	2001-3
Extra breast cancers	1,343	255
Cancers compensated in women aged 70-79 years	None (182 surplus cancers)	47
Overdiagnosed breast cancers	1,343	208
Overdiagnosis	35%	19%

Table 3: Poisson regression analysis with estimated risk ratios for in situ and invasive breast cancer combined in Denmark adjusted for increasing background incidence, geographical differences in pre-screening incidence, and delayed start of screening in Funen.

	RR (95% CI)
1971 (reference)	1.0 (APC 0,37%, 0.30-0.45)
Screening effects, women aged 50-69 years	
First round	2.07 (1.94 to 2.21)
Later rounds	1.34 (1.29 to 1.40)
Women aged 70-74 years	
1998-2003	0.89 (0.80 to 0.98)
Women aged 70-79 years	
1998-2003	0.90 (0.85-0.96)
Pre-screening difference (screened vs. rest)	
All ages	0.90 (0.89 to 0.92)

3,901 women with breast cancer would be expected, whereas we observed 5,189 cases during the 1,342,836 woman-years from 1991-2003. This yields an incidence rate ratio of 1.33, or 33% more breast cancers than expected, with a 95% confidence interval of 1.28-1.39. However, Svendsen et al. not only separated calculations of their confidence intervals for Copenhagen and Funen, they also calculated the confidence interval for each year of observation individually. Further, they left out diagnoses of CIS from their calculations. Thus, by splitting the data, Svendsen et al. only found a non-significant difference, and this formed the basis for their conclusion. We believe the study by Svendsen et al. does not provide evidence against overdiagnosis, and other authors have excluded it from their review on overdiagnosis [22].

Some of the same authors have previously claimed that, after the prevalence peak in the first rounds, the breast cancer incidence in a screened population should decline rapidly to the level expected without screening, if there were no overdiagnosis and a closed cohort was studied [23]. The Danish authors claimed that the incidence returned to the pre-screening level and that organised mammography screening could therefore operate without overdiagnosis [23]. However, they did not present a statistical analysis in support of this and have later published data for Copenhagen that are compatible with elevated incidence levels in the screened age group following the introduction of mammography screening [24].

In another study [20], some of the same authors discussed complexities in the estimation of overdiagnosis and used a lead-time model for calculating overdiagnosis in Sweden. This is an unreliable approach because current estimates of lead-time disregard overdiagnosis and its substantial influence on such calculations. Adjusting for lead-time using these estimates will therefore underestimate overdiagnosis. Most importantly, lead-time models assume that a very pronounced drop in incidence rates occurs in previously screened age groups, e.g. a drop of about 50% has been suggested, based on an assumed lead-time of 5 years [11]. However, such large drops have never occurred [10], as we have also shown here for Denmark.

Conclusions

The 33% overdiagnosis we found means that one in four breast cancers diagnosed in a screened population is overdiagnosed. This is lower than the 52% we have previously estimated for other mammography screening programmes, likely because the Danish programme has low uptake, a deliberately conservative attitude towards microcalcifications, and low recall rates. Despite these precautions, the level of overdiagnosis is still disturbingly high and it leads to overtreatment and great physical and psychological harms for those who experience it. It is therefore important that women receive balanced information that makes is possible for them to decide on a rational basis whether screening is right for them. Unfortunately, the official information leaflets women receive when they are invited to screening do not tell them about overdiagnosis and overtreatment [8]. We have therefore published an evidence-based leaflet [25] that has been translated into several languages, and which conveys the message that it is not clear whether breast screening does more good than harm [9].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KJJ and PCG conceived the project. KJJ collected data and wrote the first draft. PHZ performed statistical analyses. PHZ and PCG revised the manuscript.

Acknowledgements

No funding was needed.

References

1. Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, *et al.*: **The case for early detection.** *Nat Rev Cancer* 2003, **3:**243-52.
2. Vainio H, Bianchini F: *IARC Handbooks of Cancer Prevention. Breast Cancer Screening Volume 7*. Lyon; IARC Press; 2002.
3. Fryback DG, Stout NK, Rosenberg MA, *et al.*: **The Wisconsin breast cancer epidemiology simulation model.** *J Natl Cancer Inst Monogr* 2006, **36:**37-47.
4. Zahl PH, Mæhlen J, Welch GH: **The natural history of breast cancer detected by screening mammography.** *Arch Intern Med* 2008, **168:**2311-2316.
5. Welch GH: *Should I be tested for cancer?* Berkeley and Los Angeles, California; University of California Press, Ltd; 2004.
6. Slaytor EK, Ward JE: **How risks of breast cancer and benefits of screening are communicated to women: analysis of 58 pamphlets.** *BMJ* 1998, **317:**263-4.
7. Croft E, Barratt A, Butow P: **Information about tests for breast cancer: what are we telling people?** *J Fam Pract* 2002, **51:**858-60.
8. Jørgensen KJ, Gøtzsche PC: **Content of invitations for publicly funded screening mammography.** *BMJ* 2006, **332:**538-41.

9. Gøtzsche PC, Nielsen M: **Screening for breast cancer with mammography.** *Cochrane Database Syst Rev* 2006:CD001877.
10. Jørgensen KJ, Gøtzsche PC: **Overdiagnosis in publicly organised mammography screening programmes: Systematic review of incidence trends.** *BMJ* 2009, **339:**b2587.
11. Boer R, Warmerdam P, de Koning H, van Oortmarssen G: **Extra incidence caused by mammographic screening.** *Lancet* 1994, **343(979):**.
12. NHS Health and Social Care Information Centre: **Breast Screening Program, England: 2004-05.** [http://www.cancerscreening.nhs.uk/breastscreen/breast-statistics-bulletin-2004-05.pdf]. ISBN: 1-84636-038-2. Accessed on October 15th 2007.
13. Ries LAG, Melbert D, Krapcho M, Mariotto A, Miller BA, Feuer EJ, Clegg L, Horner MJ, Howlader N, Eisner MP, Reichman M, Edwards BK, eds: **SEER Cancer Statistics Review, 1975-2004.** 2007 [http://seer.cancer.gov/csr/1975_2004/]. National Cancer Institute, Bethesda, MD Based on November 2006 SEER data submission, posted to the SEER web site. Acessed on October 15th 2007.
14. Statistics Denmark: [http://www.statistikbanken.dk]. Accessed on February 4, 2008.
15. Danish National Board of Health: **Early detection and treatment of breast cancer: Status report.** *Copenhagen* 1997.
16. Euler-Chelpin M, Olsen AH, Njor S, Vejborg I, Schwartz W, Lynge E: **Women's patterns of participation in mammography screening in Denmark.** *Eur J Epidemiol* 2006, **21:**203-9.
17. Svendsen AL, Olsen AH, von Euler-Chelpin M, Lynge E: **Breast Cancer Incidence after the Introduction of Mammography Screening - What Should Be Expected?** *Cancer* 2006, **106:**1883-90.
18. Jensen A, Olsen AH, von Euler-Chelpin M, Njor SH, Vejborg I, Lynge E: **Do nonattenders in mammography screening programmes seek mammography elsewhere?** *Int J Cancer* 2005, **113:**464-70.
19. 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.
20. Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH: **Complexities in estimation of overdiagnosis in breast cancer screening.** *Br J Cancer* 2008.
21. Darby SC, McGale P, Taylor CW, Peto R: **Long-term mortality from heart disease and lung cancer after radiotherapy for early breast cancer: prospective cohort study of about 300,000 women in US SEER cancer registries.** *Lancet Oncol* 2005, **6:**557-65.
22. Biesheuvel C, Barratt A, Howard K, Houssami N: **Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review.** *Lancet Oncol* 2007, **8:**1129-38.
23. Olsen AH, Jensen A, Njor S, Villadsen E, Schwartz W, Vejborg I, Lynge E: **Breast cancer incidence after the start of mammography screening in Denmark.** *Br J Cancer* 2003, **88:**362-65.
24. Törnberg S, Kemetli L, Lynge E, Olsen AH, Hofvind S, Wang H, Anttila A, Hakama M, Nyström L: **Breast cancer incidence and mortality in the Nordic capitals, 1970-1998. Trends related to mammography screening programmes.** *Acta Oncol* 2006, **45:**528-35.
25. Gøtzsche PC, Hartling OJ, Nielsen M, Jørgensen KJ, Brodersen J: **Breast Screening - The Facts. Or maybe not.** *BMJ* 2009, **338:**446-8.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1472-6874/9/36/prepub