

# Mammography screening

## Benefits, harms, and informed choice

Karsten Juhl Jørgensen

This review has been accepted as a thesis together with 5 previously published papers by University of Copenhagen 20<sup>th</sup> September 2012 and defended on 11<sup>th</sup> January 2013

Official opponents: Torben Jørgensen and Anders Foldspang

Correspondence: The Nordic Cochrane Centre, Department 7811, Rigshospitalet, Blegdamsvej 9, Denmark

E-mail: kj@cochrane.dk

Dan Med J 2013;60(4): B4614

This thesis is based on the following papers:

- I. 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.
- II. Jørgensen KJ, Gøtzsche PC. Content of invitations for publicly funded screening mammography. *BMJ* 2006;332:538-541.
- III. Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark. A comparative study. *BMJ* 2010;340:c1241.
- IV. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;339:b2587.
- V. Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Women's Health* 2009;9:36.

### 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. Overdiag-

nosis 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.

### 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 recom-

mentations, 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 literature is clearly reflected in the information offered to those invited [15-18]. 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 weighed against each other, which cannot be done scientifically. It is a value judgment 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 [28-30]. 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 [39-41].

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] (Figure 1).

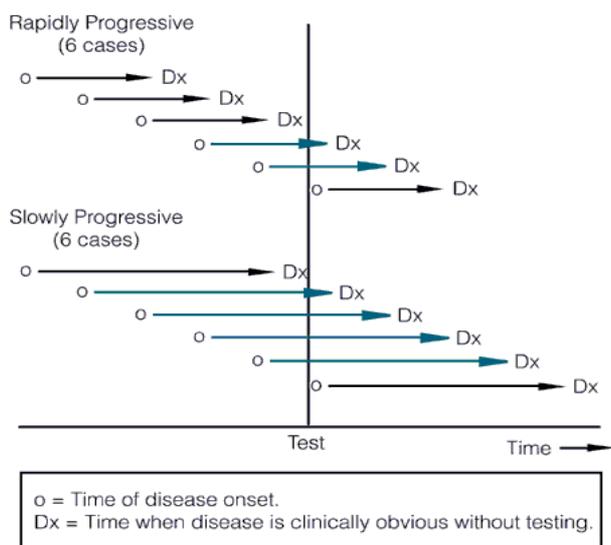


Figure 1: Length bias. From Welch 2004 [42].

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 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 (Figure 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 (Figure 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 overestimate 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 be an underestimate due to overdiagnosed small cancers. The difference in the trials corresponds to one volume doubling (Figure 2) [22].

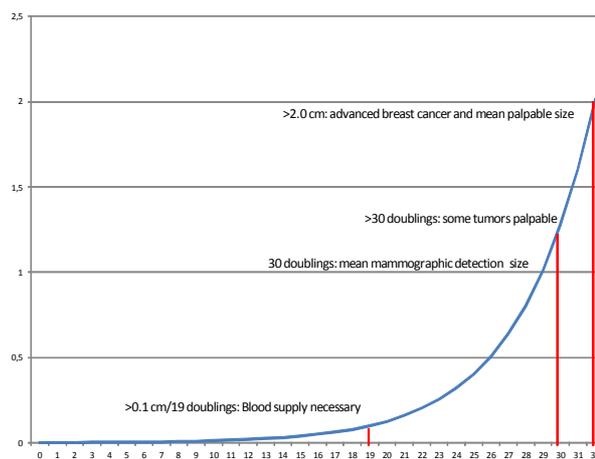


Figure 2: Tumour diameter (cm) versus volume doublings.

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 (Figure 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 consid-

erably 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 was 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 from 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 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."* [13]

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 four 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:

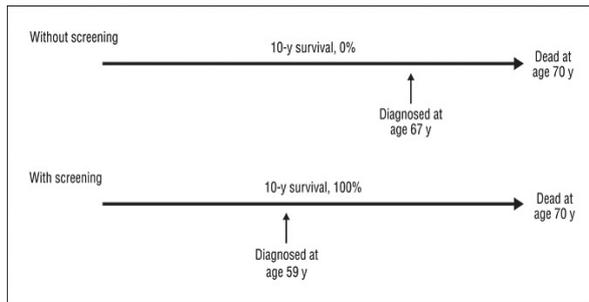
*"The screenees are the healthy, well-educated, affluent, physically fit, fruit and vegetable eating, non-smokers, with long-lived parents."* [1]

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 (Figure 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 (Figure 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

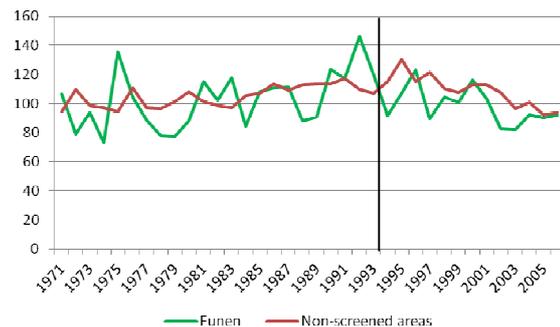
In 2005, a study reported a 25% reduction in breast cancer mortality in Copenhagen compared to unscreened regions in Denmark and a 37% reduction in breast cancer mortality among those who accepted the invitation to screening [83]. The study drew headlines such as “Cancer screening saves lives” in large 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 have 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 implementation 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 caught 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 (Figure 4).



**Figure 4:** Breast cancer mortality per 100,000 women in Funen vs. non-screened areas in Denmark. Vertical line indicates 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-screened 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]

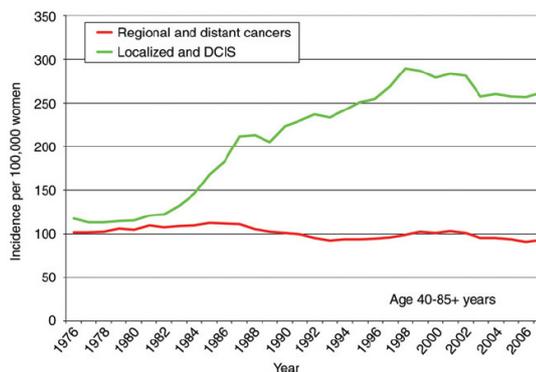
It was clear from both our study [23], and from a review of 30 European countries [96], that the effect of breast screening is too small to meet original expectations. The review found that the median change in breast cancer mortality was -37% (range -76% to -14%) in women under 50 years, -21% (-40% 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 (Figure 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 (Figure 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 neighboring 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 to 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 (Figure 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 (Figure 6) [106,107].

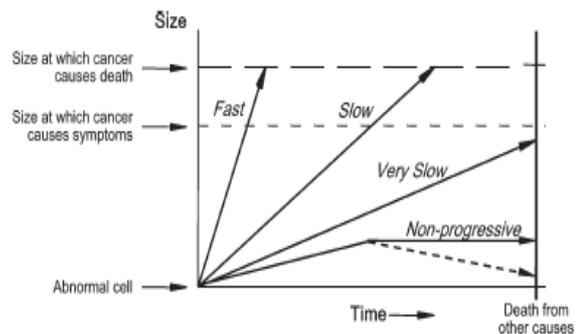


Figure 6: Variation in the growth of breast cancer. From Welch et al. 2010 [112].

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].

### 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, these 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

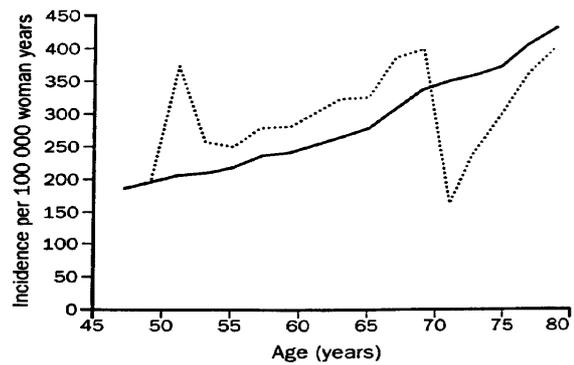
Estimating overdiagnosis in a clinical setting is important, but difficult. Often, lead-time models have been used [133], but they have important problems, as we have pointed out [134,135]. The basic premise of lead-time models is that screening causes a “shift” in the age-specific incidence due to the 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 (Figure 2), observed 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 (Figure 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. 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.



**Figure: Expected breast cancer incidence in 2-year age categories**

Solid line = not screened, dotted line = screened.

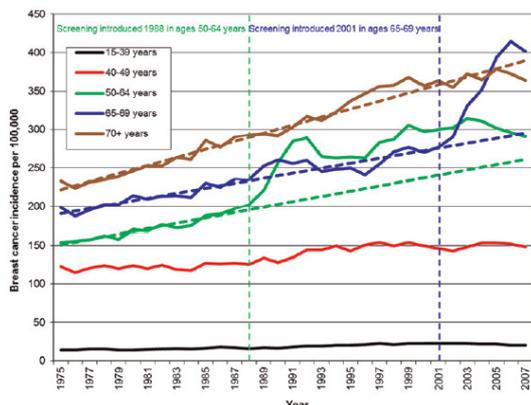
**Figure 7: Model predicting breast cancer incidence with mammography screening. From Boer et al. 1994 [139].**

If this excess incidence is not compensated by a decline in women who pass the upper 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 Figure 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 (e.g. HRT use and demographic factors leading to a higher increase in background incidence rates than projected) were likely unimportant [26].

#### 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 microcalcifications 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 overdiagnosis 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 over-treatment 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

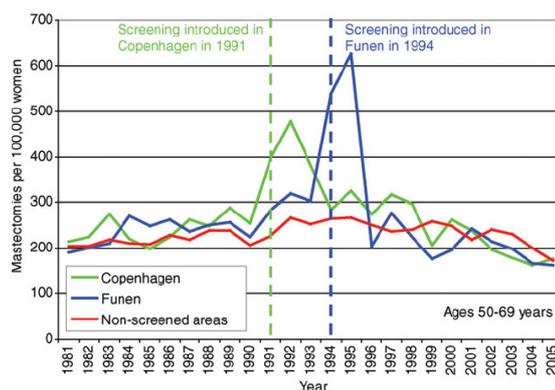
In November 2011, a letter was published in the *Lancet* by 41 signatories, all of whom had some relation to breast screening [151]. It was a reaction to the recent criticism of breast screening and focused specifically on the research from the Nordic Cochrane Centre. As we noted in our reply [152], 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 (Figure 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].



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

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].

### 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] (Figure 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 [16-18,163,164]

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.

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.



Figure 10: Campaign poster from the American Cancer Society.

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, a 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](http://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.<sup>1</sup> 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 Pro-

grammes 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 (Figure 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].

Saving lives  
through screening



**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 equalising 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 (QALYs) [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 QALYs, 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 affilia-

tions 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 [190-194] 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 acknowledges 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 (Figure 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 (Figure 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-mammo-grams.html?\\_r=1&scp=5&sq=editorial%2026%20october%202011&st=cse](http://www.nytimes.com/2011/10/27/opinion/more-questions-about-mammo-grams.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](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, Lyng 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, Vilapinyo 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. Sørli 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, Farquhar 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, Björnelid 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, Grøntoft 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* 2005;142:306-12.
  63. Moss L, Cuzick J, Clarke R, et al. 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, Pernerger 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 University 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 Malmö 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 Report 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, Nordenskjöld 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](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 Mortality in randomised trials of cancer screening. *J Natl 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. Overdiagnosis: 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 Study. *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/draftref3.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/1/hi/today/newsid\\_8596000/8596225.stm](http://news.bbc.co.uk/1/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 overdiagnosis 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 overdiagnosis 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. Luxembourg: Office 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, Carroll 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 woman with 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](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](http://www.im.dk/artikler_im_dk/Files/Fil1/3857.pdf). Accessed 21 May 2010.
159. [The Norwegian Health Legislation]. Available at: <http://www.lovdato.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, et al. Are women getting relevant information about mammography screening for an informed consent: a critical appraisal 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 Screening. *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 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](http://www.timesonline.co.uk/tol/life_and_style/health/article5776804.ece).
169. Mathieu E, Barratt A, Davey HM et al. 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 et 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 et al. 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, Chorooglou M. Possible net harms of breast cancer screening: updated modelling of Forrest report. *BMJ* 2011; 343:d7627.
178. Ioannidis JP, Evans SJ, Gøtzsche PC et al.: Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
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;290:921-8.
180. Jørgensen KJ, Johansen HK, Gøtzsche PC. Flaws in design, analysis and 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](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](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/](http://info.cancerresearchuk.org/publicpolicy/ourpolicypositions/symptom_Awareness/cancer_screening/breast-screening-review/). Accessed 6. January 2012.
196. Pisano ED, Gatsonis C, Hendrick E et al. Digital Mammographic Imaging Trial (DMIST) Investigators Group. *N Engl J 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 Cancer 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](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.