Why mammography screening has not lived up to expectations from the randomised trials.

ARTICLE *in* CANCER CAUSES AND CONTROL · NOVEMBER 2011

Impact Factor: 2.96 · DOI: 10.1007/s10552-011-9867-8 · Source: PubMed

4 AUTHORS, INCLUDING:

Karsten Juhl Jørgensen
The Nordic Cochrane Centre
89 PUBLICATIONS **1,251** CITATIONS
SEE PROFILE

Per-Henrik Zahl
Norwegian Institute of Public Health
60 PUBLICATIONS **912** CITATIONS
SEE PROFILE

Cancer Causes Control (2012) 23:15–21
DOI 10.1007/s10552-011-9867-8

REVIEW ARTICLE

Why mammography screening has not lived up to expectations from the randomised trials

Peter C. Gøtzsche · Karsten Juhl Jørgensen · Per-Henrik Zahl · Jan Mæhlen

Received: 6 April 2011 / Accepted: 28 October 2011 / Published online: 10 November 2011

Abstract We analysed the relation between tumour sizes and stages and the reported effects on breast cancer mortality with and without screening in trials and observational studies. The average tumour sizes in all the trials suggest only a 12% reduction in breast cancer mortality, which agrees with the 10% reported in the most reliable trials. Recent studies of tumour sizes and tumour stages show that screening has not lowered the rate of advanced cancers. In agreement with this, recent observational studies of breast cancer mortality have failed to find an effect of screening. In contrast, screening leads to serious harms in healthy women through overdiagnosis with subsequent overtreatment and false-positive mammograms. We suggest that the rationale for breast screening be urgently reassessed by policy-makers. The observed decline in breast cancer mortality in many countries seems to be caused by improved adjuvant therapy and breast cancer awareness, not screening. We also believe it is more important to reduce the incidence of cancer than to detect it 'early.' Avoiding getting screening mammograms reduces the risk of becoming a breast cancer patient by one-third.

Keyword Tumour size · Tumour stage · Mammography screening · Overdiagnosis · Breast cancer mortality

The expectations from mammography screening have been high. Based on the two earliest randomised trials, the New York HIP trial and the Two-County trial, the UK Forrest report assumed in 1986 that screening would lower breast cancer mortality by 30% [1]. This was supported by a meta-analysis of the four Swedish trials in 1993 that reported a 29% reduction in the age group 50–69 years [2]. In the UK population, a 25% reduction in breast cancer mortality was expected by the year 2000 [3, 4], and the American Cancer Society expected a 50% reduction by 2015 [5].

However, most of the randomised trials are of poor quality [6, 7], and comprehensive systematic reviews performed by The Nordic Cochrane Centre and the US Preventive Services Task Force have suggested that the effect was only a 15–16% reduction in breast cancer mortality [7–9].

Observational studies from 2010 have made it clear that even this modest estimate has been too optimistic. The most robust study is from Denmark, which has a unique control group, as only 20% of the rather homogeneous population was screened throughout 17 years [10]. There was not any effect of screening; in fact the decline in breast cancer mortality was slightly larger in the non-screened areas than in the screened areas (2% vs. 1% per year in the relevant age groups), and it was even larger in young women than in those women who could benefit from

Peter C. Gøtzsche wrote the first draft, all authors contributed to revisions.

P. C. Gøtzsche (✉) · K. J. Jørgensen
The Nordic Cochrane Centre, Rigshospitalet, Dept. 3343, Blegdamsvej 9, 2100 Copenhagen, Denmark
e-mail: pcg@cochrane.dk

K. J. Jørgensen
e-mail: kj@cochrane.dk

P.-H. Zahl
Norwegian Institute of Public Health, PO Box 4404, Nydalen, 0405 Oslo, Norway
e-mail: Per-Henrik.Zahl@fhi.no

J. Mæhlen
Department of Pathology, Ullevål University Hospital, 0407 Oslo, Norway
e-mail: jan.mahlen@medisin.uio.no

screening (6% vs. 5% per year). A review of 30 European countries also found that the decline was largest in young age groups that had not been screened [11], and there was no visible effect of screening either in the UK, Sweden and Norway [10, 12]. In contrast, a 2005 study reported a 15% effect in USA [13], but this might have been a statistical artefact. The decline in breast cancer mortality coincided not only with widespread propagation of screening but also with increased use of adjuvant therapy, making it difficult to separate the effects of these two interventions [14]. Furthermore, the statistical models used in this study were adjusted for a strong underlying increase in the breast cancer incidence, which was caused by screening and should therefore not have been adjusted for.

The disappointing results make it of interest to elucidate why mammography screening has not lived up to expectations. As we demonstrate below, the characteristics of the identified tumours with screening suggest that only a very marginal effect can be expected, if any.

Doubling times and heterogeneity of breast cancer

Breast cancer is a heterogeneous disease, with widely varying growth rates and metastatic potential. The growth rate for the individual cancer is usually constant for long periods of time and can therefore be described as the doubling time, which is the time it takes to double the number of malignant cells. The range of doubling times for breast cancers is very large. A US study estimated that 90% of the doubling times were between 69 and 1,622 days (median 260 days) [15]. It complicates the issue, though, that spontaneous death of cancer cells is common and may increase with tumour size [16]. This means that the doubling time increases over time, as the cancer becomes large. The growth may also stop altogether [15], and many screen-detected breast cancers are harmless, as their normal fate is to regress and disappear [17].

Cancers are usually assumed to be monoclonal, which means that the cancer has developed from a single malignant cell. From here it takes about 30 volume doublings before a cancer acquires a size of 10 mm ($\sim$ one billion cells) [16, 18, 19], which is when it becomes detectable at breast screening. If we assume that the observed doubling times also are valid from initiation till the tumour becomes detectable, the average woman has harboured the cancer for 21 years before it acquires a size of 10 mm [15]. Further, screening is more likely to pick up slow-growing tumours (length bias), i.e. those that are oldest. It is therefore misleading to say that cancers are caught 'early' with screening. They are caught very late, even though our assumption was a simplification, as cancers are likely to grow faster in the beginning.

Plausible effect based on tumour sizes in the trials

We have tabulated all the tumour data we could find from the randomised trials, based on the Cochrane review on screening [9]. There are useful data on tumour size from five trials (counting the Two-County trial as one) (Table 1) [20–23].

These data refer to invasive cancers (we shall discuss carcinoma in situ below). The average tumour size was 16 mm in the screened groups and 21 mm in the control groups. Using the formula for the volume of a sphere, we can see that a cancer of 16 mm will grow to 20 mm after only one more cell division (i.e. one volume doubling). The window of opportunity for mammography screening to work is therefore about one cell division, after the malignant cells have already divided more than 30 times. For mammography screening to have more than a marginal effect, quite many cancers would need to metastasise in this little time window. This is not plausible. Some cancers never metastasise, and some grow so slowly that it does not matter that they would ultimately have metastasised, as the women would have long been dead from other causes. Most importantly, many cancers metastasise early, before they can be detected on a mammogram. A German study of 12,423 patients found a linear correlation between tumour size and the existence of one or more positive lymph nodes [24], and we calculated that tumours with a diameter of 16 mm have metastasised in 35% of the cases. This is a conservative estimate, as many metastases are overlooked. For example, an American study of 24,740 cases from the prescreening era showed that nearly 25% of the node-negative patients eventually develop distant metastasis [25].

As screening is supposed to work by detecting cancers before they have metastasised, we can convert the tumour sizes into an expected effect of screening. The German study showed that tumours of size 16 and 21 mm are node-positive in 35 and 42% of cases, respectively, whereas the American study showed a difference of only 4% in node-positive tumours for a size difference of 5 mm. The weighted average of the two studies is 5%. If we assume for simplicity that all patients with metastases will die from

Table 1 Average size of invasive breast cancers in the randomised trials

	Screened group (mm)	Control group (mm)
Two Canadian trials	16	19
Malmö trial	14	19
Stockholm trial	14	19
Two-County trial	18	25
Mean of all trials	16	21

breast cancer, and those without will not, the expected effect of screening is a relative risk of (42–5%)/42% = 0.88, or a 12% reduction in breast cancer mortality. This agrees well with what the most reliable trials have shown. After 13 years, there was a 10% reduction in breast cancer mortality [9].

This is reassuring, considering that the tumour size difference of 5 mm has been somewhat overestimated because of overdiagnosis in the screened groups [9]. Overdiagnosed tumours are generally smaller than other tumours, as they grow more slowly (length bias). Therefore, if an overdiagnosed tumour was missed at one screening session by a small margin, it would not be much larger when detected at the next screening.

Carcinoma in situ is mainly detected at screening. As these lesions are often multifocal, it is difficult to understand how they can be considered precursors to small single tumours, as is generally stated, unless we assume that the vast majority of these lesions regress. They are often treated by mastectomy [9], and treatment with antihormone therapy also reduces the risk of new tumours occurring [26]. The prophylactic surgical treatments would only be expected to lower breast cancer mortality marginally, e.g. cancer can still arise in the remaining breast.

Lead-time

The data on tumour size raise interesting questions about published estimates of lead-time, which is the average time screening brings the diagnosis forward. There are only 2.3 times as many cells in a tumour of size 21 mm as in one of size 16 mm, which means that the average lead-time for invasive cancer is only a little more than one doubling time. In the literature, estimates of lead-time have generally varied between 2 and 5 years [27, 28]. Current screening programmes appear to detect more cancers than earlier [29, 30], and estimates of mean sojourn time (which is the time the tumour is hypothetically detectable by screening, and therefore slightly longer than lead-time) in Norway varied between 4 and 9 years [31]. The estimates of lead-time cannot be reconciled with the tumour data and with the knowledge of doubling times. The median doubling time in the US study was 260 days [15], and the mean was 212 days in a study from Heidelberg [19]. As these doubling times were derived from tumour sizes in the relevant growth intervals where tumours can be detected clinically or on a mammogram, it means that the lead-time should be less than a year.

Also epidemiological data show that lead-time estimates of 2 or more years cannot be correct. With such long lead-times, the large and persistent increase in breast cancer incidence rates that occur when screening is introduced must be followed by a huge incidence drop when women are no longer screened due to advanced age [27]. But studies have shown that a drop either does not occur or is very small [29, 32, 33].

Carcinoma in situ is a special case, as most of these lesions never progress to invasive cancer [34, 35]. They can therefore be said to have an infinite lead-time. The same is true for regressing overdiagnosed cancers. We cannot calculate averages based on numbers that include infinity, and it would also be meaningless to use hypothetical truncated numbers, for example the estimated remaining life span of women with carcinoma in situ that does not progress.

For these reasons we believe it is a dubious approach to adjust statistical analyses for lead-time estimates when studying overdiagnosis. However, this is commonly done, and by using too high estimates of lead-time and statistical models, some researchers have spuriously 'adjusted away' virtually all overdiagnosis [36–40].

Plausible effect based on tumour stages in the trials

The Swedish Two-County trial was instrumental for the introduction of screening in many countries [1]. It reported an unusually large effect of screening—a 31% reduction in mortality from breast cancer—and also a 25% reduction in the rate of advanced breast cancers (stage II or more) in the group invited to screening [41]. Tabár et al. have claimed that not only the lower stages of the tumours but also their smaller size can explain the large effect [42, 43].

We believe this is not possible, and we will show why by turning our attention from the relative risks researchers usually report to risk differences. The difference in tumour size was 7 mm, which is only slightly larger than the average of 5 mm for all the trials. Further, tumour data from the Two-County trial presented in a graph show that after 10 years, 6% of women with cancers between 10 and 14 mm in size have died, whereas 11% of those with cancers between 15 and 19 mm have died [42]. Thus, a difference in size of 5 mm corresponded to a difference in breast cancer deaths of only 5%, which, moreover, is in agreement with the difference of 5% in tumours with metastases we calculated above. However, as we showed above, this corresponds to only a 12% reduction in mortality from breast cancer, and not 31% as Tabár et al. reported.

For cancers somewhat bigger, between 20 and 29 mm, no less than 34% of women in the Two-County trial had died after 10 years [42]. This is far too many, considering that there is a close linear relationship between tumour size and risk of metastasis. In Table 2, we show the percentage of women with metastases for the same tumour size

Table 2 Tumour size as a predictor of metastases and death in breast cancer

	Tumour size (mm)	Percent with metastases or who died in breast cancer
German study	12	29% with metastases
	17	36% with metastases
	24.5	46% with metastases
Two-County trial	10–14	6% died in breast cancer
	15–19	11% died in breast cancer
	20–29	34% died in breast cancer

The three tumour sizes in the German study correspond to those reported for the Two-County trial

categories as those published for the Two-County trial. As there were 39 cancers at the first screen in the Two-County trial with a tumour size between 20 and 29 mm, the sudden increase in mortality when tumour size increased by only 7.5 mm cannot be explained by random variation. This discrepancy suggests that tumour data, assessment of cause of death, or both, are not reliable in the Two-County trial.

We have previously studied the validity of the Two-County trial by using the reported mortality reductions in the various tumour stages [44], and we found that the 31% overall mortality reduction is incompatible with the reported tumour stages [45].

There are other major problems with the Two-County trial. It is not clear how many women were randomised and the reported numbers of breast cancer deaths also varies between publications, despite similar follow-up [9, 46]. These observations prompted us to compare the reported data with official Swedish statistics, and we estimated that 192 breast cancer cases and 43 breast cancer deaths seem to be missing in the main publication of the trial from 1985 [41, 47]. We found similar discrepancies in two updates of the trial [47].

Tabár et al. [48] reported in a letter in Lancet that those trials that lowered the rate of node-positive cancers also lowered the rate of breast cancer deaths. They showed their data in a table but did not perform a statistical analysis. This relationship would be expected if screening worked, but the authors only included women in the age group 40–49 years, which was curious, as the effect of screening is considered controversial in young women. We therefore studied all screened age groups [49]. Since the purpose of screening is to advance the time of diagnosis, the trials that found many cancers, compared with the unscreened control group, would be expected to have the largest effect on breast cancer mortality, but there was no such relation. For cancers in stage II+, and also for node-positive cancers, there was a significant relation in the expected direction, but the linear relation between advanced cancers and breast cancer mortality was in the wrong place on the graph. A screening effectiveness of zero (which means that the rate of node-positive cancers is the same in the screened group as in the control group) corresponded to a significant 16% reduction in breast cancer mortality (95% confidence interval 9–23% reduction). This could only happen if there is bias. Further analyses uncovered evidence that assessment of cause of death and of the number of cancers in advanced stages was biased. Since the size of the bias, 16%, was similar to the estimated effect of screening [8, 9], this result suggested that screening might be ineffective.

No decrease in advanced cancers

Before screening was introduced in USA, the age-adjusted incidence of breast cancer was rather constant. When screening spread in the 1980s, there was a sharp rise in cases of carcinoma in situ. This would be expected to be followed by a decrease in the incidence of early-stage invasive breast cancer later on, but as this also increased [50], it indicated substantial overdiagnosis of harmless invasive cancers. More recent US data have shown that the incidence of cancers with metastases has changed very little [51, 52].

The total number of disseminated cancers also increased when screening started in Norway and remained increased even 8 years later [53]. The authors drew a wrong conclusion, however, as they said that breast cancers diagnosed in the screening period had prognostically favourable tumour characteristics. The error they made is very common. They looked at the *percentage* of advanced cancers, which went down, but this is misleading, as overdiagnosed cancers have favourable tumour characteristics. If, for example, 60 of 100 cancers in a group without screening are advanced, and the only thing screening does it to overdiagnose another 30 localised cancers (which agree fairly well with data from the Malmö trial) [21], then the *percentage* of advanced cancers is $60/(100 + 30) = 46\%$ in the screened group and $60/100 = 60\%$ in the control group. Thus, although the *absolute* rate of advanced cancers was not reduced with screening, there were *relatively* fewer advanced cancers with screening. One needs to look at *total* numbers of advanced cancers, not percentages.

Screening enthusiasts from other countries have also misled their readers. The Dutch screening programme did not reduce the incidence of cancers with metastases [54], but this fact was concealed for the readers because the researchers split the cancers with metastases in two groups, those above and below 2 cm in size. This makes no sense, but what immediately catches the eye in their figure is a graph with metastasised cancers bigger than 2 cm that appears to go down over time. The splitting of the data is misused in the abstract, which only mentions the larger

cancers and notes that their incidence declined significantly (although the total number of cancers with metastases was the same before and after screening was introduced). What the study really showed was that screening is ineffective in the Netherlands and that it leads to a huge amount of overdiagnosis (as the incidence of localised cancers doubled, with no sign of a decrease). But the authors concluded in the last sentence of their abstract in contrast to their findings that, "It is evident that breast cancer screening contributes to a reduction in advanced breast cancers and breast cancer mortality" [53].

A systematic review from 2011 that included seven countries and regions found that, on average, the rates of advanced cancers, defined as those larger than 20 mm, were not affected by screening [55].

What is the effect of screening today?

Much has happened since the trials were carried out and it is not likely that screening is effective today. There are several reasons for this.

First, when screening doesn't reduce the occurrence of advanced cancers [51–55], it cannot work. This agrees with the fact that it hasn't been possible to see an effect of screening in Denmark or in other European countries [10–12]. A comparison of three pairs of neighbouring European countries that had introduced screening 10–15 years apart showed no relation between screening start and the reductions in breast cancer mortality [56].

Second, adjuvant treatment is far better today, and as its effect is largely independent of nodal status and other tumour characteristics [57], it works whether or not the cancer is detected 'early.' When the trials were done, the women didn't receive much adjuvant therapy such as chemotherapy and anti-hormonal treatment [9]. For example, there was almost no use of anti-hormonal treatment in the Two-County trial. But it is well-documented that tamoxifen is highly effective, as is polychemotherapy.

Third, greater breast cancer awareness seems to have played a role. Breast cancer awareness is different from regular breast self-exams, as it means being aware of changes in the breasts, e.g. detected during a shower, and consulting with a doctor without delay when such changes are noted. In Denmark, the average size of the tumours was 33 mm in 1978–1979, which decreased to 24 mm in 1988–1989 [58]. This change occurred before screening started, and it is therefore a doubtful argument when screening advocates say screening has led to greater breast cancer awareness. Breast cancer awareness increased substantially *before* screening started and the Danish data show that it did not lead to overdiagnosis of breast cancer [33]. The difference of 9 mm is much greater than the average difference between the screened and the control groups in the trials, which was only 5 mm. This suggests that breast cancer awareness has been more important than screening for the decrease in breast cancer mortality that has been observed in many countries, although the contributions of breast cancer awareness has likely been small compared to the contribution of improved adjuvant therapy [57].

Fourth, there are substantial problems with reading mammograms. In the randomised trials, those who read the films were highly motivated, as they hoped to show screening worked. It is difficult in everyday practice to maintain similar concentration and carefulness, also considering that by far most of the readings are negative, in contrast to mammograms performed on a clinical suspicion of breast cancer. It is so easy to overlook the occasional cancer; in fact, it is regarded as one of the most difficult tasks in radiology to spot cancers on mammograms. Radiologists miss many breast cancers and the agreement when two or more observers evaluate the same mammograms independently is poor [59–62]. Surveys in USA have shown that average doctors missed more than 25% of the cancers, and that some clinics missed nearly 40% [63]. These problems would tend to reduce the effect of screening compared to the trials.

Fifth, the amount of overdiagnosis seems to have increased since the trials were performed. In New South Wales [30], researchers found 42% overdiagnosis of invasive breast cancer with an interpolation method, and 30% with an extrapolation method when they used a lead-time of 5 years, and higher percentages, 51% and 36%, when the lead-time was 2.5 years. We found 38% overdiagnosis of invasive breast cancer in the same age group in New South Wales and 53% when we included carcinoma in situ [29]. The authors reported that the detection rates had increased substantially within the programme itself, from about 28 to 42 per 10,000 women screened [30]. As these increased detection rates were not accompanied by decreased interval cancer rates [30], one would not expect screening to reduce breast cancer mortality, and the authors suggested that the increased detection was harmful, as it simply produced more overdiagnosis. With more overdiagnosis comes increased mortality from the harms of radiotherapy and chemotherapy given to healthy women.

Conclusions

Screening has not delivered the expected effect, and there are two reasons for this. First, the expectation of a 30% reduction in breast cancer mortality was based on poor trials. The likely effect of screening at the time the randomised trials were performed was somewhere between no

effect and a 12% reduction in breast cancer mortality. We base this on our analyses of tumour sizes and stages; on the reported effects in the most reliable trials [9]; and on the bias in assessing cause of death and whether cancers are node-positive [9, 49].

Second, recent observational studies and studies of tumour sizes and stages indicate that screening may not have an effect today. Overdiagnosis, which results in overtreatment, has a small prophylactic effect on breast cancer mortality, but it leads to increased mortality from other causes, and the net effect of screening must be marginal, if any.

In contrast, screening leads to serious harms in healthy women through overdiagnosis and false-positive findings [9]. We therefore suggest that the rationale for breast screening be urgently reassessed by policy-makers. There might of course be an occasional woman who might benefit from screening, but that is not a good argument for screening. This could be said about almost any intervention. Perhaps it is useful to compare with drugs. Many drugs are withdrawn from the market, although they benefit many patients, when serious harms are reported in rather few patients. The situation with mammography screening is the reverse of this: Very few will benefit, if any, whereas many will be harmed. We also believe it is more important to reduce the incidence of cancer than to detect it 'early.' The US Center for Medical Consumers has noted that if we wish to reduce the incidence of breast cancer, there is nothing as effective as avoiding getting screening mammograms [64]. With an overdiagnosis of 52% [29], avoiding screening mammograms reduces the risk of becoming a breast cancer patient by one-third.

Conflict of interest None.

References

1. Forrest P (ed) (1986) Breast cancer screening. Report to the health ministers of England, Wales, Scotland & Northern Ireland. Department of Health and Social Science
2. Nyström L, Rutqvist LE, Wall S et al (1993) Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet 341:973–978
3. Austoker J (1994) Screening and self examination for breast cancer. BMJ 309:168–174
4. Department of Health (1992) The health of the nation: a strategy for health in England. HMSO, London
5. Byers T, Mouchawar J, Marks J et al (1999) The American Cancer Society challenge goals. How far can cancer rates decline in the U.S. by the year 2015? Cancer 86:715–727
6. Gøtzsche PC, Olsen O (2000) Is screening for breast cancer with mammography justifiable? Lancet 355:129–134
7. Olsen O, Gøtzsche PC (2001) Screening for breast cancer with mammography (Cochrane Review). In: The Cochrane Library, Issue 4. Update Software, Oxford
8. Humphrey LL, Helfand M, Chan BK, Woolf SH (2002) Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 137(5 Part 1):347–360
9. Gøtzsche PC, Nielsen M (2009) Screening for breast cancer with mammography. Cochrane Database Syst Rev (4). Art. No.: CD001877
10. Jørgensen KJ, Zahl P-H, Gøtzsche PC (2010) Breast cancer mortality in organised mammography screening in Denmark: comparative study. BMJ 340:c1241
11. Autier P, Boniol M, LaVecchia C et al (2010) Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. BMJ 341:c3620
12. Kalager M, Zelen M, Langmark F, Adami HO (2010) Effect of screening mammography on breast-cancer mortality in Norway. N Engl J Med 363:1203–1210
13. Berry DA, Cronin KA, Plevritis SK et al (2005) Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 353:1784–1792
14. Correspondences and author reply (2006) Screening and breast cancer. N Engl J Med 354:767–769
15. Spratt JS, Meyer JS, Spratt JA (1996) Rates of growth of human neoplasms: part II. J Surg Oncol 61:68–83
16. Friberg S (2005) On the growth rates of human malignant tumors: implications for medical decision making. J Oncol 55:1–22
17. Zahl PH, Gøtzsche PC, Mæhlen J (2011) Natural history of breast cancers detected in the Swedish mammography screening program; a cohort study. Lancet Oncol. doi:10.1016/S1470-2045(11)70250-9
18. Skrabanek P (1985) False premises and false promises of breast cancer screening. Lancet 2:316–320
19. Fournier Dv, Weber E, Hoeffken W et al (1980) Growth rate of 147 mammary carcinomas. Cancer 45:2198–2207
20. Narod SA (1997) On being the right size: a reappraisal of mammography trials in Canada and Sweden. Lancet 349:1869
21. Andersson I, Aspegren K, Janzon L et al (1988) Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. BMJ 297:943–948
22. Frisell J, Eklund G, Hellstrom L, Somell A (1987) Analysis of interval breast carcinomas in a randomized screening trial in Stockholm. Breast Cancer Res Treat 9:219–225
23. Tabar L, Fagerberg G, Chen HH et al (1996) Tumour development, histology and grade of breast cancers: prognosis and progression. Int J Cancer 66:413–419
24. Engel J, Eckel R, Kerr J et al (2003) The process of metastasisation for breast cancer. Eur J Cancer 39:1794–1806
25. Carter CL, Allen C, Henson DE (1989) Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer 63:181–187
26. Cuzick J, Sestak I, Pinder SE et al (2011) Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: long-term results from the UK/ANZ DCIS trial. Lancet Oncol 12:21–29
27. Boer R, Warmerdam P, de Koning H, van Oortmarssen G (1994) Extra incidence caused by mammographic screening. Lancet 343:979
28. Tabár L, Duffy SW, Yen M-F et al (2002) All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. J Med Screen 9:159–162
29. Jørgensen KJ, Gøtzsche PC (2009) Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ 339:b2587
30. Morrell S, Barratt A, Irwig L et al (2010) Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. Cancer Causes Control 21:275–282

31. Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S (2008) Estimating mean sojourn time and screening sensitivity using questionnaire data on time since previous screening. J Med Screen 15:83–90
32. Zahl PH, Strand BH, Maehlen J (2004) Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. BMJ 328:921–924
33. Jørgensen KJ, Zahl P-H, Gøtzsche PC (2009) Overdiagnosis in organised mammography screening in Denmark: a comparative study. BMC Women's Health 9:36
34. Nielsen M, Thomsen JL, Primdahl S et al (1987) Breast cancer atypia among young and middle-aged women: a study of 110 medicolegal autopsies. Br J Cancer 56:814–819
35. Welch HG, Black WC (1997) Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast. Ann Intern Med 127:1023–1028
36. Yen MF, Tabar L, Vitak B et al (2003) Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer 39:1746–1754
37. Smith RA, Duffy SW, Gabe R et al (2004) The randomized trials of breast cancer screening: what have we learned? Radiol Clin N Am 42:793–806
38. Paci E, Warwick J, Falini P, Duffy SW (2004) Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? J Med Screen 11:23–27
39. Duffy SW, Agbaje O, Tabar L et al (2005) Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. Breast Cancer Res 7:258–265
40. Olsen AH, Agbaje OF, Myles JP et al (2006) Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. Breast J 12:338–342
41. Tabar L, Fagerberg CJG, Gad A et al (1985) 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 1:829–832
42. Tabar L, Fagerberg G, Duffy SW et al (1992) Update of the Swedish two-county program of mammographic screening for breast cancer. Radiol Clin N Am 30:187–210
43. Tabar L, Fagerberg CJG, Day NE (1988) The results of periodic one-view mammographic screening in Sweden. Part 2: evaluation of the results. Hans Huber, Toronto, pp 39–44
44. Tabar L, Fagerberg G, Chen HH et al (1995) Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer 75:2507–2517
45. Zahl PH, Kopjar B, Mæhlen J (2001) Norske tall for brystkreftdødelighet og validititet i svenske mammografistudier. Tidsskr Nor Laegeforen 121:1928–1931
46. Gøtzsche PC (2004) On the benefits and harms of screening for breast cancer. Int J Epidemiol 33:56–64
47. Zahl P-H, Gøtzsche PC, Andersen JM, Mæhlen J (2006) Results of the Two-County trial of mammography screening are not compatible with contemporaneous official Swedish breast cancer statistics. Dan Med Bull 53:438–440
48. Tabár L, Smith RA, Duffy SW (2002) Update on effects of screening mammography. Lancet 360:337
49. Gøtzsche PC (2011) Relation between breast cancer mortality and screening effectiveness: systematic review of the mammography trials. Dan Med Bull 58:A4246
50. Fletcher SW, Elmore JG (2003) Mammographic screening for breast cancer. N Engl J Med 348:1672–1680
51. Esserman L, Shieh Y, Thompson I (2009) Rethinking screening for breast cancer and prostate cancer. JAMA 302:1685–1692
52. Jørgensen KJ, Keen JD, Gøtzsche PC (2011) Is mammographic screening justifiable considering its substantial overdiagnosis rate and minor effect on mortality? Radiology 260:621–626
53. Hofvind S, Sorum R, Thoresen S (2008) Incidence and tumor characteristics of breast cancer diagnosed before and after implementation of a population-based screening-program. Acta Oncol 47:225–231
54. Fracheboud J, Otto SJ, van Dijck JA, Broeders MJ, Verbeek AL, de Koning HJ (2004) Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. Br J Cancer 91:861–867
55. Autier P, Boniol M, Middleton R et al (2011) Advanced breast cancer incidence following population based mammographic screening. Ann Oncol 20 Jan (Epub ahead of print)
56. Autier P, Boniol M, Gavin A, Vatten LJ (2011) Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. BMJ 343:d4411
57. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687–1717
58. Rostgaard K, Vaeth M, Rootzén H, Lynge E (2010) Why did the breast cancer lymph node status distribution improve in Denmark in the pre-mammography screening period of 1978–1994? Acta Oncol 49:313–321
59. Kerlikowske K, Grady D, Barclay J et al (1998) Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Natl Cancer Inst 90:1801–1809
60. Beam CA, Layde PM, Sullivan DC (1996) Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample. Arch Intern Med 156:209–213
61. Poplack SP, Tosteson AN, Grove MR et al (2000) Mammography in 53,803 women from the New Hampshire mammography network. Radiology 217:832–840
62. Sickles EA, Wolverton DE, Dee KE (2002) Performance parameters for screening and diagnostic mammography: specialist and general radiologists. Radiology 224:861–869
63. Moss M (2002) Spotting breast cancer: doctors are weak link. New York Times 27 June
64. Napoli M (2009) Reduce your risk of breast cancer: avoid mammograms. July 14, 2009. http://medicalconsumers.org/2009/07/14/cut-your-risk-of-breast-cancer-avoid-mammograms/