

RESEARCH

Overdiagnosis in screening mammography in Denmark: population based cohort study

 OPEN ACCESS

Sisse Helle Njor *statistician*¹, Anne Helene Olsen *statistician*², Mogens Blichert-Toft *professor emeritus*³, Walter Schwartz *chief physician*⁴, Ilse Vejborg *chief physician*⁵, Elsebeth Lynge *professor*¹

¹Department of Public Health, University of Copenhagen, Østre Farimagsgade 5, DK 1014 Copenhagen K, Denmark; ²Institute of Community Medicine, University of Tromsø, Tromsø, Norway; ³Danish Breast Cancer Cooperative Group, 2100 Copenhagen Ø, Denmark; ⁴Mammography Screening Clinic, University Hospital Odense, 5000 Odense, Denmark; ⁵Diagnostic Centre, University Hospital Copenhagen, Blegdamsvej, 2100 Copenhagen Ø, Denmark

Abstract

Objective To use data from two longstanding, population based screening programmes to study overdiagnosis in screening mammography.

Design Population based cohort study.

Setting Copenhagen municipality (from 1991) and Funen County (from 1993), Denmark.

Participants 57 763 women targeted by organised screening, aged 56-69 when the screening programmes started, and followed up to 2009.

Main outcome measures Overdiagnosis of breast cancer in women targeted by screening, assessed by relative risks compared with historical control groups from screening regions, national control groups from non-screening regions, and historical national control groups.

Results In total, 3279 invasive breast carcinomas and ductal carcinomas in situ occurred. The start of screening led to prevalence peaks in breast cancer incidence: relative risk 2.06 (95% confidence interval 1.64 to 2.59) for Copenhagen and 1.84 (1.46 to 2.32) for Funen. During subsequent screening rounds, relative risks were slightly above unity: 1.04 (0.85 to 1.27) for Copenhagen and 1.14 (0.98 to 1.32) for Funen. A compensatory dip was seen after the end of invitation to screening: relative risk 0.80 (0.65 to 0.98) for Copenhagen and 0.67 (0.55 to 0.81) for Funen during the first four years. The relative risk of breast cancer accumulated over the entire follow-up period was 1.06 (0.90 to 1.25) for Copenhagen and 1.01 (0.93 to 1.10) for Funen. Relative risks for participants corrected for selection bias were estimated to be 1.08 for Copenhagen and 1.02 for Funen; for participants followed for at least eight years after the end of screening, they were 1.05 and 1.01. A pooled estimate gave 1.040 (0.99 to 1.09) for all targeted women and 1.023 (0.97 to 1.08) for targeted women followed for at least eight years after the end of screening.

Conclusions On the basis of combined data from the two screening programmes, this study indicated that overdiagnosis most likely amounted to 2.3% (95% confidence interval -3% to 8%) in targeted women. Among participants, it was most likely 1-5%. At least eight years after the end of screening were needed to compensate for the excess incidence during screening.

Introduction

The purpose of screening mammography is to reduce mortality from breast cancer without increasing mortality from other diseases. Preventive measures in healthcare might, however, also have unintended negative side effects, and the occurrence of these should be closely monitored. In screening mammography, the most serious concern is the risk of overdiagnosis—that is, diagnosis of breast cancer that would in the absence of screening not have led to clinically manifest disease in the woman's lifetime.¹ Overdiagnosis cannot be identified biologically, as distinguishing between progressive and non-progressive or slowly progressive cancers is not possible with current diagnostic tools. Overdiagnosis can therefore be investigated only epidemiologically.

Screening affects the incidence rate. Assuming a three year advancing of time of diagnosis (lead time) and screening of all women during a two year period, a doubling of the incidence rate is expected during the first round of screening.² As screening continues, the incidence rate should go down to the level before screening, apart from an increase caused by the artificial aging—that is, breast cancer diagnosed at age 55 in the absence of screening will during screening be diagnosed, for example, at age 52. A complementary dip in the incidence rate is expected after women leave the screening programme.^{3 4} Overdiagnosis occurs if the cumulative incidence some years after the end of screening is higher than the cumulative incidence expected in

Correspondence to: S H Njor sissenj@sund.ku.dk

Extra material supplied by the author (see <http://www.bmj.com/content/346/bmj.f1064?tab=related#webextra>)

the absence of screening.⁵ This pattern may be further complicated by possible temporal changes in the underlying incidence rate.

In most randomised controlled trials, screening was offered to the control groups after the end of the trials, making estimation of the amount of overdiagnosis difficult. The Malmö trial in Sweden left the control group unscreened. For women screened at age 45-79 and followed past screening age, this randomised controlled trial showed a relative cumulative incidence rate of 1.10 (95% confidence interval 0.99 to 1.22) for the intervention compared with the control group.⁶ The control groups were not offered screening in the Canadian randomised controlled trials either,^{7,8} but as these trials targeted only women aged 40-49 and 50-59, participants probably continued screening elsewhere. With the limited data from randomised controlled trials, data on overdiagnosis from service screening programmes are warranted. The challenge in observational studies is to measure the expected cumulative incidence rate of breast cancer in the absence of screening. By using both current and historical data from the screening regions, as well as current and historical data from non-screening regions, we can measure the change in incidence of breast cancer from before to during screening in the screening regions controlled for the underlying temporal trend in breast cancer incidence. Following this method, we present here estimates of overdiagnosis from Denmark, where population based, service screening mammography was implemented in two regions in the early 1990s while nationwide roll-out of screening took place in the late 2000s.

Material and methods

Screening organisation

The population based screening mammography programme in Copenhagen started on 1 April 1991, inviting women aged 50-69 at the start of each biennial invitation round. The participation rate was 71% in the first round. The programme in Funen started on 1 November 1993 and invited women aged 50-69 at the date of invitation. The participation rate was 84% in the first round. See web appendix for details.

Use of opportunistic screening has always been limited in Denmark. In 2000 only 3% of women aged 50-69 in non-screening regions had a mammogram taken; this included diagnostic as well as opportunistic screening mammograms.⁹ The nationwide roll-out of screening mammography started only in 2007.

Data

The Danish central population register holds information on current and historical addresses for all citizens. From this register, we identified women targeted by screening in Copenhagen from 1 April 1991 to 31 March 2005 and in Funen from 1 November 1993 to 31 October 2004. To be able to detect a possible compensatory dip above the age of invitation to screening, a sufficiently long follow-up time should be included in the analysis. From Copenhagen, we therefore included only women born 1 April 1921 to 31 March 1935—that is, women aged 56-70 years on 1 April 1991 (start of screening in Copenhagen); from Funen, we included women born 1 November 1923 to 31 October 1934—that is, women aged 59-70 on 1 November 1993 (start of screening in Funen) (fig 1). For each of the two screening regions, we constructed three control groups: a historical control group, a national control group from non-screening regions, and a historical national control group (table 1).

The historical control groups included birth cohorts of women from the two respective screening regions from before the screening started. We followed these birth cohorts in the same age range as the birth cohorts invited to screening (table 1). The national control group included the same birth cohorts as the respective study groups but from non-screening regions. The historical national control groups were constructed in the same way as the historical control groups. The nationwide roll-out of screening targeting women aged 50-69 started in 2007, so this nationwide programme did not affect women born before 1937.

For the study group, we defined the entry date as the first date the woman was present in the target age group and screening region. We used similar definitions for the control groups. Person years at risk were accumulated from the woman's entry into the respective study or control group until diagnosis of invasive breast carcinoma or ductal carcinoma in situ, emigration, death, or end of follow-up on 31 December 2009, whichever came first. We identified incident cases of invasive breast carcinoma and ductal carcinoma in situ by linkage to the Danish cancer register and the clinical database of the Danish Breast Cooperative Group. Linkage was based on personal identification numbers. We defined participants as targeted women who attended screening at least once. Data on participation came from the mammography screening registers in Funen and Copenhagen. The mammography screening registers hold individual information on all participants.

Analysis

We estimated overdiagnosis as the cumulative incidence of breast cancer in a screening region compared with the expected cumulative incidence in the absence of screening. If the incidence of breast cancer in the absence of screening had developed equally in screening and non-screening regions, the expected incidence in the absence of screening could be estimated from the incidence in the historical control group controlled for the change in incidence from historical to present time in the national control group. Our study therefore has three control groups: the historical control group, the national control group, and the historical national control group.

However, as we included data on incidence of breast cancer for a period of almost 33 years, from 1977 to 2009 in Copenhagen, the background incidence in the absence of screening might have developed differently in the screening region and in the non-screening region (that is, interaction between region and period). We therefore started out by testing the assumption of no interaction between region and period. As separating the interaction between period and region from the effect of screening is not possible, we used data from the five year period before screening. Studying the interaction in the same cohorts as included in the study and control groups is important, as the interactions might vary by cohorts. We therefore used the same cohorts as in the study and control groups but looked at the five year period before the study period and control periods. This means that we used data on incidence of breast cancer in the screening cohorts for a pre-study period of five years before the first invitation to screening: the period 1 April 1986 to 31 March 1991 for Copenhagen and 1 November 1988 to 31 October 1993 for Funen. We used equivalent pre-study periods for the three control groups (table 1). We used Poisson regression for this analysis (model A in appendix).

For Funen, we found no statistically significant interaction between region and period for any cohort ($P=0.22-0.93$), and we saw no trends. We consequently analysed the Funen data

for the study period by using a Poisson model without interaction between region and period (model B in appendix). For Copenhagen, we found a statistically significant interaction between region and period for the four oldest cohorts ($P=0.046$), as illustrated in the appendix figure; the incidence of breast cancer in the Copenhagen study group in the pre-study period developed differently from the incidence in the control groups. We consequently took this interaction into account in the analysis of the Copenhagen data, assuming that for each two year cohort the interaction in the study period equalled the interaction found in the pre-study period, and using data from both the pre-study and the study periods (model C in appendix).

We analysed data separately for five time periods in the study group: programme prevalence screening round, programme incidence screening rounds, 0-3 years after end of screening, 4-7 years after end of screening, and at least 8 years after end of screening. As the numbers of women followed for more than 12 years were relatively small, we did not divide the analysis into the time periods 8-11 years, 12-15 years, and 16 years and above. We present the results as relative risks with two sided 95% confidence intervals. To get an estimate of overdiagnosis based on data from both Copenhagen and Funen, we made a pooled estimate by using a fixed effects weighted average of the relative risks on a logarithmic scale.¹⁰

Results

In the Copenhagen study group, we included 32 931 women and followed them for an average of 13.9 years, giving 456 499 person years; 1892 invasive breast carcinomas and 110 cases of ductal carcinoma in situ occurred (table 2). The population of Copenhagen was considerably larger during the historical control period, including 63 097 women followed for an average of 14.4 years, giving 909 875 person years; 2598 invasive breast carcinomas and 41 cases of ductal carcinoma in situ occurred. The national control group and the historical national control group for Copenhagen were fairly equal in size, including 281 311 women followed for an average of 14.8 years and 266 860 women followed for an average of 15.0 years, giving 4 173 549 and 3 999 172 person years. The numbers of invasive breast carcinomas differed, being 14 410 in the national control group and 10 323 in the historical national control group, a difference reflecting the underlying increase in incidence of breast cancer.¹¹

In the Funen study group, we included 24 832 women and followed them for an average of 13.0 years, giving 323 363 person years; 1203 invasive breast carcinomas and 74 cases of ductal carcinoma in situ occurred (table 3). The size of the Funen population was fairly constant over time, with 27 143 women followed for an average of 13.2 years in the historical control period, giving 359 426 person years; 1040 invasive breast carcinomas and 45 cases of ductal carcinoma in situ occurred. The national control group and the historical national control group for Funen were also of fairly equal size, including 213 380 and 209 443 women, both followed for an average of 13.0 years, giving 2 768 352 and 2 731 477 person years and 9898 and 7635 invasive breast carcinomas.

Copenhagen

The cumulative incidence in Copenhagen during screening was 5% higher than expected in the absence of screening (relative risk 1.05, 95% confidence interval 0.88 to 1.24) (fig 2). Inclusion of ductal carcinoma in situ changed the estimate only slightly, giving a relative risk of 1.06 (0.90 to 1.25). If 1000 targeted women were followed over 20 years, the relative risk of 1.06 would translate into 87 invasive breast carcinomas plus

ductal carcinomas in situ compared with 82 cases expected in the absence of screening. We observed a doubling of the incidence of breast cancer during the programme prevalence peak (relative risk 2.06, 1.64 to 2.59; invasive breast carcinoma plus ductal carcinoma in situ) (fig 2). The incidence was close to the expected incidence in the absence of screening during the programme incidence screening rounds (relative risk 1.04, 0.85 to 1.27). We saw a clear deficit during the first 0-3 years after the end of screening (relative risk 0.80, 0.65 to 0.98), after which the incidence gradually approached the level expected in the absence of screening. We found no significant difference in relative risk for cumulative incidence by age at entry (heterogeneity test $P=0.47$).

Funen

The cumulative incidence in Funen was 1% higher than expected (relative risk 1.01, 0.92 to 1.10) (fig 2). Inclusion of ductal carcinoma in situ had a marginal effect on this estimate (relative risk 1.01, 0.93 to 1.10). If 1000 targeted women were followed over 20 years, the relative risk of 1.01 would translate into 78 invasive breast carcinomas plus ductal carcinomas in situ compared with 77 cases expected in the absence of screening. The incidence of breast cancer almost doubled during the programme prevalence peak (relative risk 1.84, 1.46 to 2.32; invasive breast carcinoma plus ductal carcinoma in situ) (fig 2) and was non-significantly increased during the programme incidence screening rounds (1.14, 0.98 to 1.32). We saw a clear deficit during the first 0-3 years after end of screening (relative risk 0.67, 0.55 to 0.81), after which the incidence gradually approached the level expected in the absence of screening. We found no significant difference in relative risk for cumulative incidence by age at entry (heterogeneity test $P=0.97$).

At least eight years of follow-up

Including only birth cohorts with at least eight years of follow-up after the end of screening (Copenhagen 1921-31; Funen 1923-31), and including both invasive breast carcinoma and ductal carcinoma in situ, gave a relative cumulative incidence of relative risk 1.034 (0.86 to 1.25) for Copenhagen and 1.007 (0.91 to 1.12) for Funen.

Pooled estimates

The pooled estimate for all targeted women was 1.040 (0.99 to 1.09). The pooled estimate for women with at least eight years of follow-up was 1.023 (0.97 to 1.08).

Estimates for participants

From the mammography screening register, we know who has ever participated in one of the screening programmes. In Copenhagen, 32% of the studied birth cohorts had never participated in screening. Compared with the expected level in the absence of the screening programme, these women had a relative cumulative incidence (invasive breast carcinoma plus ductal carcinoma in situ) of 0.88. When we used these results to control for selection bias among screening participants, the relative cumulative incidence in participants became 1.08 (see appendix for details on methods). In Funen, 29% of the studied birth cohorts had never participated in screening, and the relative cumulative incidence for these women was 0.96, giving a relative cumulative incidence for participants of 1.02. In the birth cohorts followed for eight years or more after the end of screening, the relative cumulative incidence for participants became 1.05 in Copenhagen and 1.01 in Funen.

Discussion

Our data indicate that among women targeted for screening, the excess lifetime risk of invasive breast carcinoma plus ductal carcinoma in situ amounted to 6% (95% confidence interval –10% to 25%) in the Copenhagen screening programme and 1% (–7% to 10%) in the Funen programme. Among targeted women followed for at least eight years after the end of screening, the excess risk amounted to 3% (–14% to 25%) in the Copenhagen programme and 0.7% (–9% to 12%) in the Funen programme. Among participants, the excess risk amounted to 1-5% for participants followed for at least eight years after the end of screening.

In these Danish screening mammography programmes, the incidence of breast cancer over time developed in accordance with the introduction of a lead time. The incidence increased with the start of screening, was somewhat above the level without screening in the subsequent screening rounds, and decreased significantly below this level once invitation to screening had stopped. The relative risk for incidence of invasive breast carcinoma plus ductal carcinoma in situ was still below unity 4-7 years after the end of screening (0.91, 0.73 to 1.14, for Copenhagen and 0.78, 0.64 to 0.94, for Funen), suggesting that the compensatory drop after the end of screening took at least eight years. Therefore, the most reliable estimate of overdiagnosis should be based on women followed for at least eight years after the end of screening and should include both invasive breast carcinoma and ductal carcinoma in situ.

If 1000 women targeted by screening in Copenhagen were followed for a total of 20 years, of which at least eight years were after end of screening, the relative risk of 1.034 would translate into 86.9 cases of invasive breast carcinoma and ductal carcinoma in situ in the screening group compared with 84.0 cases expected in the absence of screening. For Funen, a similar calculation would give 78.3 observed cases compared with 77.7 cases expected in the absence of screening.

In the Danish screening programmes, in which ductal carcinoma in situ in the first three invitation rounds constituted 13-14% of screen detected cases,^{12 13} the excess breast cancer risk in targeted women changed only slightly after inclusion of ductal carcinoma in situ. This was because cases of ductal carcinoma in situ constituted only a minor part of all the cases in the study group. This finding stresses the seriousness of any overdiagnosis, as invasive cancer is treated more intensively than ductal carcinoma in situ.

The relatively modest estimates of overdiagnosis found in this study make good sense, considering that screening is carried out in middle aged women in developed countries and that these women have an excellent life expectancy, leaving considerable time for preclinical disease to progress to symptoms in the absence of screening. Although occasional documented cases of regression of untreated tumours occur, untreated breast cancer is generally characterised by very low survival.¹⁴

Strengths and weaknesses of study

To our knowledge, this is the first analysis of overdiagnosis in screening mammography based on individual follow-up of all women targeted by screening as well as of women in population based control groups. Follow-up of women targeted by screening has also been reported from Florence, Italy.^{15 16} Our study was based on register data, which were all linkable owing to the use of the personal identification numbers as the common identifier in population and health registers. Our analysis took the underlying increase in incidence of breast cancer into account.

Our study was entirely observational without assumptions on transition probabilities or length of lead time. As pointed out by de Gelder et al,¹⁷ the choice of denominator plays an important role in the calculation of overdiagnosis. To be able to inform women about the long term consequences of participation in screening, we calculated the cumulative incidence of invasive breast carcinoma plus ductal carcinoma in situ after a first invitation to screening divided by the cumulative incidence of invasive breast carcinoma plus ductal carcinoma in situ expected during the same lifespan in the absence of screening.

A limitation in the “three control group” design was the lack of control of possible interaction between region and period. We compensated for this by analysing our study group and three control groups in a pre-study period. The interaction between region and period was statistically significant for Copenhagen and was included in the analysis, whereas this was not the case for Funen. The difference between the two regions was most likely explained by the decreasing size and possibly changed composition of the Copenhagen population over time and the inclusion of historical control birth cohorts back to 1907 in Copenhagen but only to 1912 in Funen. Eventual earlier uptake of hormone therapy use in Copenhagen could also have contributed to the interaction. The Copenhagen study population was larger than the study population from Funen; however, the inclusion of the interaction term in the analysis for Copenhagen resulted in broader confidence intervals than for Funen. For Copenhagen, we estimated the interaction in the study period by assuming that it equalled the interaction found in the pre-study period. In view of the increase over time in the pre-study period in the incidence for the Copenhagen cohorts born in 1921-25 (see appendix figure), we might have overestimated the overdiagnosis in Copenhagen. Inclusion of an interaction term in the analysis for Funen only marginally changed the results.

Another limitation was inclusion only of women first targeted at ages 56-69 for Copenhagen and 59-69 for Funen, to ensure follow-up time after end of screening. Results from this study therefore pertain to women aged 56 and above at first invitation. These women had, however, considerable screening experience as they had been invited up to seven times in Copenhagen and six times in Funen. Puliti et al found the excess incidence of breast cancer in the last year of screening to be quite similar for women aged 50-54, 55-59, and 60-64 years at the start of service screening.¹⁶ We therefore expect our estimate of overdiagnosis to be fairly representative also for women aged 50-69 at the start of service screening. Finally, the overdiagnosis in screening is expected to be affected by the detection rates for ductal carcinoma in situ. On the basis of crude proportion data, detection of ductal carcinoma in situ varies across screening settings, constituting 13-14% of screen detected cases in the two Danish programmes compared with 24% in US screening data.¹⁸ Overdiagnosis therefore has to be considered in view of the actual screening context.

Comparison with other studies

A wide range of estimates of overdiagnosis have been published. Here, we focus on estimates based on data from Denmark and neighbouring Nordic countries. In our earlier analysis of breast cancer incidence rates for women aged 50-69 in Copenhagen and Funen in the first years after the introduction of screening,¹⁹ we found marked prevalence peaks (consistent with the “red areas” in this study) and incidence rates within the expected time trend during subsequent screening rounds (consistent with the “yellow areas”). Using Markov modelling, Olsen et al

estimated overdiagnosis based on data from the first two rounds of the Copenhagen programme to be 7.8% (0.3% to 26.5%) for the first round and 0.5% (0.02% to 2.1%) for the second round.²⁰ Given the broad confidence intervals, these results were in agreement with those of our study. Jørgensen et al analysed the time trends in incidence of invasive breast carcinoma plus ductal carcinoma in situ for the period 1971–2003.²¹ This study merged data from Copenhagen, Funen, and Frederiksberg and used age groups instead of birth cohorts, population data instead of individual data, and a short follow-up period. The authors concluded that screening led to a 33% overdiagnosis. This crude estimate was not compatible with the outcome from the long term birth cohort data presented here.

Zahl et al compared the cumulative incidence of breast cancer in cohorts of Norwegian women offered screening three times with cohorts offered screening once and found a relative risk of 1.22 (1.16 to 1.30).²² Using the same study design, Zahl et al found a relative risk of 1.14 (1.10 to 1.18) for cohorts of Swedish women offered screening three times compared with cohorts offered screening once.²³ These relative risks are to some extent comparable to our results from the subsequent screening rounds (“yellow areas”), and the estimated relative risk of 1.14 for Sweden was in fact the same as we found for Funen. The higher relative risk found in Norway has been linked to changes in use of hormone therapy.²⁴

Dutch simulation data illustrated the importance of length of follow-up in studies of overdiagnosis.¹⁷ The same was seen in our data, where the relative risk was below unity not only 0–3 years after the end of screening but also 4–7 years after. If the follow-up had ended four years after end of screening, our study would have indicated an overdiagnosis of 9–10% instead of 1–6%. The importance of length of follow-up has been shown previously with data from the Malmö randomised trial, in which Moss found an overdiagnosis of 31% when the trial was still ongoing,²⁵ whereas Zackrisson et al found an overdiagnosis of 8–10% when they included follow up to 15 years after the trial.⁶

Conclusions and policy implications

This cohort study based on data from two longstanding population based screening mammography programmes indicated that overdiagnosis of breast cancer including ductal carcinoma in situ most likely amounted to 2.3% (95% confidence interval –3% to 8%) in women aged 56–69 years when first targeted by screening. Among participating women, overdiagnosis of breast cancer including ductal carcinoma in situ most likely amounted to 1% in Funen and 5% in Copenhagen. Our study furthermore illustrated that at least eight years after the end of screening were needed to compensate for the excess incidence during screening.

Contributors: SHN participated in the design of the study, did the statistical analysis, participated in the analysis and interpretation of data, and drafted the manuscript. AHO participated in the design of the study, retrieved and checked data, and revised the manuscript critically. MB-T, WS, and IV contributed data on screening mammography and revised the manuscript critically. EL conceived the study, participated in the design of the study, and finalised the manuscript in collaboration with SHN. All authors have read and approved the final manuscript. All authors had full access to all of the data in the study. SHN is the guarantor.

Funding: This study was financially supported by the Esper and Olga Boel Foundation. The funding source had no role in the writing of the manuscript or the decision to submit it for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: SHN and EL had financial support from Esper and Olga Boel Foundation; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: In accordance with Danish legislation, the study was notified to the Danish Data Inspection Agency (J No 2008-41-2191). No further ethical approval is needed for register based studies in Denmark.

Data sharing: No additional data available.

- Paci E, Duffy S. Overdiagnosis and overtreatment in service screening. *Breast Cancer Res* 2005;7:266-70.
- Duffy SW, Lyng E, Jonsson H, Ayyaz S, Olsen AH. Complexities in the estimation of overdiagnosis in breast cancer screening. *Br J Cancer* 2008;99:1176-8.
- Boer R, Warmerdam P, de Koning H, van Oortmarssen G. Extra incidence caused by mammographic screening. *Lancet* 1994;343:979.
- Møller B, Weedon-Fekjaer H, Hakulinen T, Tryggvadóttir L, Storm HH, Talbäck M, et al. The influence of mammographic screening on national trends in breast cancer incidence. *Eur J Cancer Prev* 2005;14:117-28.
- 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.
- Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* 2006;332:689-92.
- 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. *Ann Intern Med* 2002;137:E305-8.
- Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study—2: 13-year results of a randomised trial in women aged 50-59 years. *J Natl Cancer Inst* 2000;92:1490-9.
- Jensen A, Olsen AH, von Euler-Chelpin M, Njor SH, Vejborg I, Lyng E. Do nonattenders in mammography screening programmes seek mammography elsewhere? *Int J Cancer* 2005;113:464-70.
- Harris RJ, Bradburn MJ, Deeks JJ, Harbord RM, Altman DG, Sterne JAC. Meta: fixed and random effects meta-analysis. *Stat J* 2008;8:3-28.
- Tryggvadóttir L, Gislum M, Bray F, Klint A, Hakulinen T, Storm HH, et al. Trends in the survival of patients diagnosed with breast cancer in the Nordic countries 1964-2003 followed up to the end of 2006. *Acta Oncol* 2010;49:624-31.
- Lyng E. Mammography screening for breast cancer in Copenhagen April 1991-March 1997. *APMIS Suppl* 1998;83:1-44.
- Njor SH, Olsen AH, Bellström T, Dyrberg U, Bak M, Axelsson C, et al. Mammography screening in the county of Fyn November 1993-December 1999. *APMIS Suppl* 2003;110:1-33.
- Johnstone PA, Norton MS, Riffenburgh RH. Survival of patients with untreated breast cancer. *J Surg Oncol* 2000;73:273-7.
- Puliti D, Miccinesi G, Zappa M, Manneschi G, Crocetti E, Paci E. Balancing harms and benefits of service mammography screening programs: a cohort study. *Breast Cancer Res* 2012;14:R9.
- Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* 2009;45:3166-71.
- De Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011;33:111-21.
- National Cancer Institute, Breast Cancer Surveillance Consortium. Cancers for 2,264,089 screening mammography examinations from 2002 - 2006—based on BCSC data as of 2009. <http://breastscreening.cancer.gov/data/benchmarkscreening/2009/table4.html>.
- Svendsen AL, Olsen AH, von Euler-Chelpin M, Lyng E. Breast cancer incidence after the introduction of mammography screening: what should be expected? *Cancer* 2006;106:1883-90.
- Olsen AH, Agbaje OF, Myles JP, Lyng E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J* 2006;12:338-42.
- Jørgensen KJ, Zahl PH, Gøtzsche PC. Overdiagnosis in organised mammography screening in Denmark: a comparative study. *BMC Womens Health* 2009;9:36.
- Zahl PH, Mæhlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 2008;168:2311-6.
- Zahl PH, Gøtzsche PC, Mæhlen J. Natural history of breast cancers detected in the Swedish mammography screening programme: a cohort study. *Lancet Oncol* 2011;12:1118-24.
- Weedon-Fekjaer H, Sørum R, Brenn MK. Hormone therapy use may explain recent results regarding tumor regression. *Arch Intern Med* 2009;169:996-7.
- Moss S. Overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Res* 2005;7:230-4.

Accepted: 05 February 2013

Cite this as: *BMJ* 2013;346:f1064

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-commercial License, which permits use, distribution, and reproduction in any medium, provided the original work is properly cited, the use is non commercial and is otherwise in compliance with the license. See: <http://creativecommons.org/licenses/by-nc/2.0/> and <http://creativecommons.org/licenses/by-nc/2.0/legalcode>.

What is already known on this topic

- Overdiagnosis in screening mammography is a widely debated topic
- Most studies have major methodological limitations
- Cohort studies of women followed past their screening age are warranted

What this study adds

- Population based data on overdiagnosis were controlled for the underlying trend in incidence of breast cancer
- Overdiagnosis amounted to 2.3% in screen targeted women and to 1-5% in participants
- At least eight years after end of screening were needed to compensate for the excess incidence during screening

Tables

Table 1 | Definition of study and control groups

	Study group	Historical control group*	National control group†	Historical national control group‡
Copenhagen				
Birth cohorts	1 Apr 1921-31 Mar 1935	1 Apr 1907-31 Mar 1921	1 Apr 1921-31 Mar 1935	1 Apr 1907-31 Mar 1921
Pre-study period	1 Apr 1986-31 Mar 1991	1 Apr 1972-31 Mar 1977	1 Apr 1986-31 Mar 1991	1 Apr 1972-31 Mar 1977
Study period	1 Apr 1991-31 Dec 2009	1 Apr 1977-31 Dec 1995	1 Apr 1991-31 Dec 2009	1 Apr 1977-31 Dec 1995
Funen				
Birth cohorts	1 Nov 1923-31 Oct 1934	1 Nov 1912-31 Oct. 1923	1 Nov 1923-31 Oct 1934	1 Nov 1912-31 Oct 1923
Pre-study period	1 Nov 1988-31 Oct 1993	1 Nov 1977-31 Oct 1982	1 Nov 1988-31 Oct 1993	1 Nov 1977-31 Oct 1982
Study period	1 Nov 1993-31 Dec 2009	1 Nov 1982-31 Dec 1998	1 Nov 1993-31 Dec 2009	1 Nov 1982-31 Dec 1998

*Included women born 14 years before women in study group; this means that no women in historical control group have been invited to screening.

†Included women living in parts of Denmark where no screening was offered; these women belonged to same birth cohorts as women in study group.

‡Included women born 14 years before women in national control group; these women belonged to same birth cohorts as women in historical control group.

Table 2| Breast cancer cases, ductal carcinoma in situ (DCIS) cases, and person years during follow-up from 1991 to 2009 of screening mammography programme in Copenhagen, Denmark, and for three control groups

Study group	Pre-study (0)	Prevalence screening (1)	Incidence screening (2)	0-3 years after screening (3)	4-7 years after screening (4)	≥8 years* after screening (5)	Prevalence screening to 4 years after screening	Prevalence screening to 8 years after screening	Total (1-5)
Study group									
Breast cancer†	339	365	619	364	296	248	1348	1644	1892
DCIS	14	39	50	9	9	3	98	107	110
Person years‡	134 400	61 423	168 592	100 531	72 912	53 042	330 546	403 457	456 499
Breast cancer rate	252.2	594.2	367.2	362.1	406.0	467.6	407.8	407.5	414.5
Breast cancer + DCIS rate	262.0	650.6	393.7	360.3§	414.3	467.5§	431.4	428.3	432.9
Historical control group									
Breast cancer†	465	295	900	638	449	316	1833	2282	2598
DCIS	4	3	10	13	9	6	26	35	41
Person years‡	263 353	119 663	347 102	198 165	143 470	101 475	664 930	808 400	909 875
Breast cancer rate	176.6	246.5	259.3	322.0	313.0	311.4	275.7	282.3	285.5
Breast cancer + DCIS rate	178.2	249.1	262.3	328.7	319.5	317.6	279.7	286.8	290.2
National control group									
Breast cancer†	2383	1520	5248	3292	2516	1834	10 060	12 576	14 410
DCIS	77	38	170	75	66	50	283	349	399
Person years‡	1 187 390	548 833	1 608 296	896 881	654 197	465 342	3 054 010	3 708 207	4 173 549
Breast cancer rate	200.7	277.0	326.3	367.1	384.6	394.1	329.4	339.1	345.3
Breast cancer + DCIS rate	206.8	283.2	335.0	373.3	393.7	403.5	336.9	347.0	353.3
Historical national control group									
Breast cancer†	1566	1201	3561	2386	1844	1331	7148	8992	10 323
DCIS	13	11	43	51	34	20	105	139	159
Person years‡	1 054 444	523 929	1 571 234	858 175	615 204	430 630	2 953 337	3 568 542	3 999 172
Breast cancer rate	148.5	229.2	226.6	278.0	299.7	309.1	242.0	252.0	258.1
Breast cancer + DCIS rate	149.8	231.4	229.3	283.5	304.9	312.6	245.4	255.6	261.8

*Mean follow-up for women with more than eight years of follow-up after end of screening=11.8 years.

†Numbers of breast cancers included in calculation of breast cancer rates; when combined breast cancer and DCIS rates were calculated, some of these cases were excluded because women had previously been diagnosed as having DCIS and were consequently censored at that point in time.

‡Person years for calculation of breast cancer rates; owing to censoring at first breast cancer or DCIS, person years for calculation of combined breast cancer and DCIS rates were slightly lower and are not reported here.

§Incidence rates for breast cancer + DCIS were lower than incidence rates for breast cancer alone because 12 women diagnosed as having breast cancer in period 0-3 years after screening had been diagnosed as having DCIS during screening, and four women diagnosed as having breast cancer in period ≥8 after screening had been diagnosed as having DCIS in an earlier period.

Table 3| Breast cancer cases, ductal carcinoma in situ (DCIS) cases, and person years during follow-up from 1993 to 2009 of screening mammography programme in Funen, Denmark, and for three control groups

Study group	Pre-study (0)	Prevalence screening (1)	Incidence screening (2)	0-3 years after screening (3)	4-7 years after screening (4)	≥8 years* after screening (5)	Prevalence screening to 4 years after screening	Prevalence screening to 8 years after screening	Total (1-5)
Study group									
Breast cancer†	248	255	436	207	212	93	898	1110	1203
DCIS	23	26	31	6	7	4	63	70	74
Person years‡	103 907	42 739	114 556	80 710	63 870	21 488	238 004	301 875	323 363
Breast cancer rate	238.7	596.7	380.6	256.5	331.9	432.8	377.3	367.7	372.0
Breast cancer + DCIS rate	260.1	659.4	401.5	260.2	339.8	453.2	400.0	387.2	391.6
Historical control group									
Breast cancer†	205	117	322	277	245	79	716	961	1040
DCIS	3	4	19	13	6	3	36	42	45
Person years‡	118 488	47 929	129 100	90 427	69 665	22 305	267 455	337 120	359 426
Breast cancer rate	173.0	244.1	249.4	306.3	351.7	354.2	267.7	285.1	289.4
Breast cancer + DCIS rate	175.0	252.6	263.7	316.8	359.5	368.3	279.6	296.1	300.6
National control group									
Breast cancer†	2085	1165	3353	2513	2122	745	7031	9153	9898
DCIS	80	34	101	75	48	24	210	258	282
Person years‡	909 958	372 669	988 289	688 017	539 065	180 311	2 048 975	2 588 041	2 768 352
Breast cancer rate	229.1	312.6	339.3	365.3	393.6	413.2	343.1	353.7	357.5
Breast cancer + DCIS rate	237.2	321.3	347.3	373.9	400.8	425.2	351.5	361.8	365.9
Historical national control group									
Breast cancer†	1625	821	2532	2055	1690	537	5408	7098	7635
DCIS	12	19	50	38	31	15	107	138	153
Person years‡	913 259	369 564	993 883	684 032	52 0815	163 183	2 047 479	2 568 294	2 731 477
Breast cancer rate	177.9	222.2	254.8	300.4	324.5	329.1	264.1	276.4	279.5
Breast cancer + DCIS rate	179.1	227.2	259.5	305.4	328.9	338.6	269.0	281.1	284.6

*Mean follow-up for women with more than eight years of follow-up after end of screening=9.9 years.

†Numbers of breast cancers included in calculation of breast cancer rates; when combined breast cancer and DCIS rates were calculated, some of these cases were excluded because women had previously been diagnosed as having DCIS and were consequently censored at that point in time.

‡Person years for calculation of breast cancer rates; person years for calculation of combined breast cancer and DCIS rates were slightly lower and are not reported here.

Figures

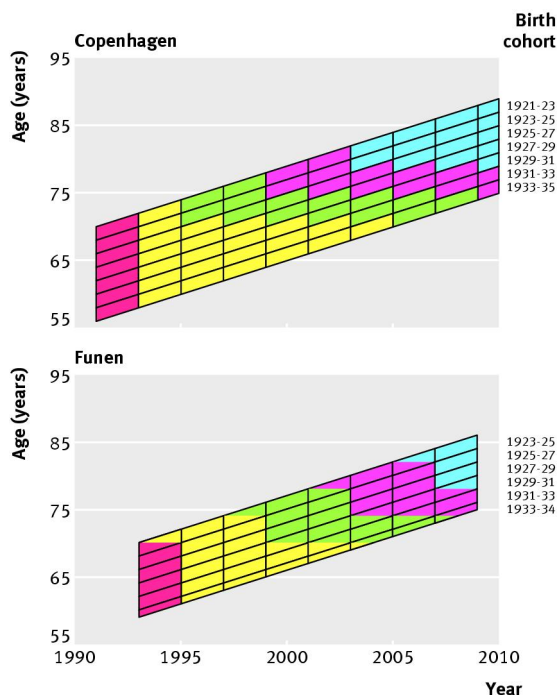


Fig 1 Study groups in Copenhagen and Funen. Red=programme prevalence screening rounds; yellow=programme incidence screening rounds; green=0-3 years after end of invitation to screening; pink=4-7 years after end of invitation to screening; blue= \geq 8 years after end of invitation to screening. Copenhagen birth cohorts: 1 April 1921-31 March 1923; 1 April 1923-31 March 1925; 1 April 1925-31 March 1927; 1 April 1927-31 March 1929; 1 April 1929-31 March 1931; 1 April 1931-31 March 1933; 1 April 1933-31 March 1935. Funen birth cohorts: 1 November 1923-31 October 1925; 1 November 1925-31 October 1927; 1 November 1927-31 October 1929; 1 November 1929-31 October 1931; 1 November 1931-31 October 1933; 1 November 1933-31 October 1934

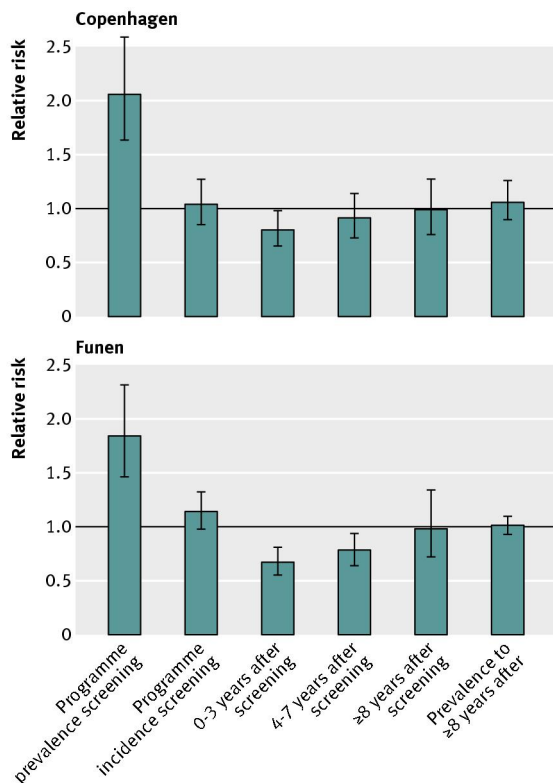


Fig 2 Relative risks and 95% confidence intervals for incidence of invasive breast carcinoma and ductal carcinoma in situ by time during and after end of invitation to screening. “Prevalence to ≥ 8 years after” means starting with prevalence round and including follow-up eight and more years after end of invitation to screening