Natural history of breast cancers detected in the Swedish mammography screening programme: a cohort study

Per-Henrik Zahl, Peter C Gøtzsche, Jan Mæhlen

Lancet Oncol 2011; 12: 1118–24

Published Online
October 12, 2011
DOI:10.1016/S1470-2045(11)70250-9

See Comment page 1083

Norwegian Institute of Public Health, Nydalen, Oslo, Norway (P-H Zahl DrMedSci); **Nordic Cochrane Centre, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark** (Prof P C Gøtzsche DrMedSci); **and Department of Pathology, University Hospital-Ullevål, Oslo, Norway** (Prof J Mæhlen DrMedSci)

Correspondence to:
Dr Per-Henrik Zahl, Norwegian Institute of Public Health, PO Box 4404 Nydalen, N-0403 Oslo, Norway
per-henrik.zahl@fhi.no

Summary

Background **The natural history of screen-detected breast cancers is not well understood. A previous analysis of the incidence change during the introduction of the Norwegian screening programme in the late 1990s suggested that the natural history of many screen-detected invasive breast cancers is to regress spontaneously but the study was possibly confounded by use of hormone replacement therapy in the population. We did a similar analysis of data collected during an earlier period when few women were exposed to hormone replacement therapy.**

Methods **We compared cumulative breast cancer incidence in age-matched cohorts of women living in seven Swedish counties before and after the initiation of public mammography screening between 1986 and 1990. Women aged 40–49 years were invited to screening every year and women aged 50–74 years were invited every 2 years. A screened group including all women aged 40–69 years (n=328 927) was followed-up for 6 years after the first invitation to the programme. A control group including all women in the same age range (n=317 404) was also followed-up for 6 years—4 years without screening and 2 years when they entered the screening programme. Screening attendance was much the same in both groups (close to 80%). Counts of incident invasive breast cancers were obtained from the Swedish Cancer Registry (in-situ cancers were excluded).**

Findings **Before the age-matched controls were invited to be screened at the end of their follow-up period, the 4-year cumulative incidence of invasive breast cancer was significantly higher in the screened group (982 per 100 000) than it was in the control group (658 per 100 000) (relative risk [RR] 1·49, 95% CI 1·41–1·58). Even after prevalence screening in the control group, the screened group had higher 6-year cumulative incidence of invasive breast cancer (1443 per 100 000 *vs* 1269 per 100 000; RR 1·14, 1·10–1·18).**

Interpretation **Because the cumulative incidence among controls did not reach that of the screened group, we believe that many invasive breast cancers detected by repeated mammography screening do not persist to be detected by screening at the end of 6 years, suggesting that the natural course of many of the screen-detected invasive breast cancers is to spontaneously regress.**

Funding **None.**

Introduction

A systematic review[1] of breast cancer incidence data in five countries (Australia, Canada, Norway, Sweden, and UK) showed that the introduction of mammography screening was associated with a 52% increase in breast cancer incidence. Because very little of this increase was compensated for by a decrease in incidence of breast cancer in previously screened women, about one in three breast cancers detected in a population offered organised screening is overdiagnosed.[1]

On the basis of an analysis of breast cancer incidence in four Norwegian counties during a 10-year period,[2] which included the start of a programme of screening every 2 years in 1996–97, we proposed that many overdiagnosed (and screen-detected) cancers would have undergone spontaneous regression if they had not been treated. In that analysis,[2] accumulated breast cancer incidence in a study group who were screened three times between 1996 and 2001 was 22% higher than it was in a control group who were screened only once at the end of the 6-year period (1992–97).

However, our methods and interpretation of the results have been criticised.[3] A major objection has been that we might have overestimated the incidence increase caused by screening, because another possible cause for the increase in breast cancer incidence in the 1990s was the increased use of hormone replacement therapy (HRT) in postmenopausal women.[3] Thus, our analysis could have been confounded by greater use of HRT—and therefore more HRT-induced breast cancers—in the screened group (followed up between 1996 and 2001) than in the control group (followed up between 1992 and 1997).

Here, we have taken account of this criticism by doing a similar analysis with a Swedish dataset that was collected during an earlier period when far fewer postmenopausal women used HRT.

Methods

Population

We collected cohort data for invasive breast cancer (in-situ carcinoma not included) from the Swedish Cancer Registry, covering 1975 to 2009. Nearly all cancers in the population-based Swedish Cancer Registry have been

proven invasive by histopathology.[4] In seven counties (Uppsala, Södermannland, Jönköping, Kalmar, Örebro, Västernorrland, and Norrbotten), about 375 000 women aged 40–74 years (about 22% of the age-eligible Swedish population) were invited to annual screening (those younger than 50 years) or screening every 2 years (those aged 50 years or older) for the first time between Aug 1, 1986, and Jan 1, 1990. Jönköping county had two screening programmes, starting on Aug 1, 1986, and April 1, 1987, respectively.[5,6]

We excluded data from the other 14 counties in Sweden. In five of them (Stockholm, Västra Götaland, Skåne, Dalarne, and Östergötland) we could not ascertain when women were invited to screening because randomised mammography screening trials were done. In four counties (Gotland, Värmland, Jämtland, and Västerbotten) screening started after 1992 and therefore coincided with the rapid increase in use of HRT. In another four counties (Blekinge, Västmanland, Halland, and Kronoborg), the screened age range was smaller than 40–74 years. Finally, we excluded Gävleborg because screening started in 1974 as part of a non-randomised study.

Design and statistical analysis

All women aged 40–74 years were invited for screening. We grouped together all women who were aged 40–69 years when they were invited to the prevalence screening (screened group). The women in the screened group were aged 45–74 years after they were invited to the third screening round (we excluded women aged 70–74 years because these women would leave the screening programme within 6 years and would therefore be invited to fewer than three screening rounds). Our control group included all women who were aged 40–69 years 4 years before they were invited to attend prevalence screening—ie, after prevalence screening these women were also aged 45–74 years.

At entry, 317 404 women were in the control group and 328 927 women were in the screened group (figure 1). 272 154 women in the screened group (83%) had, at a younger age, also been in the control group. The mean age at the beginning of the study was 55·2 years in the control group and 54·7 years in the study group.

In each county, the eldest four cohorts of women were in only the control group and the youngest four cohorts of women were in only the screened group, the other cohorts were in both the control and the screened groups. The change in number of exposure years from one year to the next in a cohort is caused by migration and by mortality (mostly for the oldest). In figure 2, which shows the number of invasive breast cancers, we have re-aligned the breast cancer incidence data from each county according to the screening start date—ie, year 1 begins at the start date of screening.

Our staggered study design has been described previously.[2] We compared the cumulative 6-year breast cancer incidence of a test group invited to multiple screenings with that of a control group invited to only one screening by the end of the 6-year follow-up period. We did this because the inclusion of a prevalence screen in the control group should give rise to a high detection rate that would compensate for the extra incidence increase seen in the test group. Our null hypothesis was that the number of breast cancers in both groups after the 6-year follow-up period would be much the same—ie, that no invasive breast cancers regress.

In-situ cancers, such as ductal carcinoma in situ, were excluded in the calculation of the cumulative incidence. Thus, the primary outcome was the cumulative incidence of invasive breast cancer during a 6-year period. The denominator for our incidence calculations—the number

					First screening		Second screening		Third screening	
	Year –3	Year –2	Year –1	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
40	**13715**	14401	14304	14289	**13880**	13597	13249	13003	12568	12413
41	12795	**13719**	14372	14301	14300	**13871**	13633	13317	13029	12601
42	12330	12762	**13717**	14328	14275	14296	**13881**	13702	13348	13029
43	11807	12285	12747	**13678**	14318	14280	14298	**13889**	13729	13329
44	11018	11758	12278	12746	**13636**	14322	14252	14290	**13868**	13724
45	10521	10977	11742	12271	12722	**13631**	14327	14267	14306	**13840**
46	10211	10501	10961	11703	12270	12690	13639	14326	14260	14272
47	9969	10188	10486	10914	11664	12259	12686	13593	14328	14256
48	9861	9941	10186	10438	10861	11659	12251	12693	13600	14272
49	9430	9830	9885	10177	10413	10861	11634	12243	12679	13594
50	9579	9415	9806	9865	10162	10362	10827	11599	12222	12667
51	9608	9555	9404	9772	9851	10142	10370	10827	11572	12196
52	9640	9595	9549	9369	9749	9841	10132	10367	10810	11556
53	9808	9598	9542	9508	9356	9731	9796	10117	10373	10776
54	9799	9763	9582	9494	9469	9325	9672	9777	10085	10346
55	10173	9741	9722	9565	9468	9446	9308	9650	9784	10065
56	10093	10143	9689	9691	9521	9451	9423	9284	9635	9750
57	10249	10046	10137	9649	9646	9495	9421	9410	9294	9635
58	10425	10217	10004	10086	9626	9627	9452	9397	9379	9277
59	10359	10366	10191	9973	10038	9578	9635	9426	9371	9351
60	10551	10329	10319	10135	9939	9986	9532	9621	9378	9327
61	11120	10460	10287	10247	10089	9892	9925	9516	9572	9329
62	11477	11041	10415	10219	10170	10020	9862	9841	9465	9542
63	11063	11394	10973	10334	10144	10094	9974	9799	9783	9420
64	10624	10970	11316	10890	10255	10086	10061	9917	9748	9706
65	10821	10546	10877	11211	10780	10158	10021	9987	9874	9716
66	10513	10723	10474	10757	11084	10685	10063	9979	9909	9811
67	9828	10405	10591	10365	10662	10974	10552	9975	9908	9803
68	10073	9693	10237	10456	10259	10522	10850	10444	9863	9811
69	9944	9902	9548	10098	10320	10125	10405	10725	10318	9748
70	9800	9782	9747	9377	9927	10181	9978	10255	10559	10175
71	9588	9616	9605	9599	9203	9756	10008	9816	10103	10398
72	9619	9365	9411	9438	9399	9043	9582	9799	9634	9960
73	9285	9424	9130	9230	9236	9210	8842	9380	9627	9404
74	8975	9035	9169	8873	9012	9019	8965	8640	9169	9421

Figure 1: **Number of woman-years at risk, by age and year**
The control group is purple and green; the study group is green and red. The first cohorts in the control and study groups are in bold. Year 1 is the start of screening. Examples of corresponding cohorts in the screened and the control group (bold numbers) can be followed diagonally.

					First screening		Second screening		Third screening	
	Year –3	Year –2	Year –1	Year 0	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6
40	**8**	7	2	15	**12**	9	16	9	6	9
41	5	**13**	16	11	22	**18**	15	18	15	11
42	14	13	**9**	12	19	16	**17**	13	22	10
43	9	8	10	**14**	20	19	20	**16**	10	16
44	17	8	17	13	**17**	25	23	14	**24**	24
45	15	11	17	14	14	**24**	25	20	29	**21**
46	10	16	18	17	24	26	18	21	15	31
47	17	20	12	16	22	27	23	23	22	21
48	9	7	17	21	13	22	23	21	30	28
49	10	9	9	16	21	22	20	22	16	32
50	14	15	10	13	23	28	22	21	22	25
51	11	12	13	8	18	18	23	22	24	26
52	16	10	25	8	12	27	21	13	26	19
53	14	15	16	14	25	28	17	17	16	22
54	13	17	18	16	25	30	30	21	18	20
55	9	9	9	17	25	24	24	11	18	20
56	19	16	16	10	28	17	24	20	26	18
57	16	20	19	15	25	25	27	23	21	24
58	21	15	12	27	39	23	19	14	25	20
59	19	15	22	23	29	29	30	24	22	25
60	14	17	15	16	22	19	31	22	17	21
61	25	22	21	16	34	32	24	26	18	24
62	11	25	25	27	34	35	23	23	20	28
63	18	26	21	23	23	37	37	24	24	14
64	26	21	28	12	42	37	35	28	25	26
65	21	25	27	31	38	39	24	31	24	35
66	29	24	25	16	40	65	34	31	24	30
67	22	24	21	25	37	44	39	33	23	23
68	16	18	18	34	40	36	37	29	28	37
69	33	20	19	26	50	52	49	38	29	36
70	27	29	19	22	55	59	27	30	32	38
71	29	35	29	23	44	44	34	35	25	43
72	17	34	27	20	49	51	37	40	30	34
73	20	26	29	23	35	41	28	22	36	22
74	14	27	30	24	23	35	20	16	22	14

Figure 2: **Number of invasive breast cancers, by age and year**
The control group is purple and green; the study group is green and red. The first cohorts in the control and study groups are in bold. Year 1 is the start of screening. Examples of corresponding cohorts in the screened and the control group (bold numbers) can be followed diagonally.

See Online for webappendix

of women in each 1-year age group for each year of our analysis—were obtained from Statistics Sweden.[7]

We calculated the annual incidence of invasive breast cancer in the screened and control groups and compared the cumulative incidence in women in the screened group with age- matched women in the control group at years 4, 6, 8, and 10. We then calculated the relative risk (RR; screened *vs* control) for being diagnosed with invasive breast cancer during the follow-up period by dividing the cumulative incidence in the screened group by that in the control group. Because data from the first 2 years of the screening programme contributed to both the screened and the control group in the 6-year analysis, we empirically estimated 95% CIs using the bootstrap technique[8] (simulating the distribution of rates of all cells), which allows the numerator and denominator of the RR to be dependent variables. We estimated the underlying incidence rate in the 10-year period before screening started using a Poisson regression model, adjusting only for 5-year age groups (with use of datasets for groups A and B in the webappendix). We used STATA (version 10) and Gauss (version 3.6) for statistical analysis.

To study how the results were affected by age, we calculated RRs for 5-year age groups (40–44 years, 41–45 years, 42–46 years, and so on through age 65–69 years). In these calculations, only about 20% of the women were in both groups. The emerging RRs represent a sort moving average, but they have another purpose. In our analysis of individuals aged 40–69 years at study entry, a substantial proportion of women were in both the screened and control group (although at different times in their life—ie, a 40-year-old women at entry in the control group would be 44 years old at entry in the screened group; about 83% are individually matched in the age group 40–69 as opposed to 20% in the 5-year age groups). This analysis was done to determine whether individual matching had an effect on the estimated RR—ie, to determine whether the differences in individual risk variables other than age could explain our results.

Role of the funding source

There was no funding source for this study. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

The start of the screening programme from 1986 to 1990 was associated with a substantial increase in breast cancer incidence (figure 3) and the attendance rate was about 80%.[6] After realignment of incidence data from each county so that year 1 starts with the starting date of the screening programme, the incidence peak caused by the prevalence screen is more pronounced (66%) and the incidence increase after the prevalence peak is about 35% (figure 3). We also recorded an annual change in breast cancer incidence before screening started of 0·7% (95% CI –0·2 to 1·6), but this increase was not significant (p=0·11; figure 3).

After 4 years of follow-up, the cumulative incidence was 982 per 100 000 in the screened group and 658 per 100 000 in the control group; the absolute difference in cumulative rates was 324 per 100 000 (RR 1·49, 95% CI 1·41–1·58; figure 4). After 6 years of follow-up, the cumulative incidence was 1443 per 100 000 in the screened group and 1269 per 100 000 in the control group; the absolute difference was 174 per 100 000 (RR 1·14, 1·10–1·18; figure 4). This finding suggests that more than half (174 of 324 per 100 000) of the incidence increase in the first two screening rounds are cancers that in the absence of these two screenings would not have not been detected in the third screening round. During the next 4 years of follow-up, the incidences in the two groups were much the same; the absolute difference between groups was 147 per 100 000 after 8 years of follow-up and 181 per 100 000 after

10 years of follow-up. The average RR of all 5-year age groups in the study was 1·16 (95% CI 1·12–1·20; table 1).

In women aged 40–47 years, the average incidence was stable at about 100 per 100 000 in the 10-year period before screening started. In the first 6 years of the screening programme, in which these women were invited annually, the recorded incidence per 100 000 was 143, 158, 149, 131, 139, and 150 in years 1–6, respectively (linear trend analysis: RR 0·98, 95% CI 0·94–1·04; p=0·40). Because women aged 40–49 years were invited yearly, we could also use our study design to compare cumulative 3-year incidence in two groups of women aged 40–47 years; one test group invited each year and one control group invited in the third year only. 3-year cumulative breast cancer incidence was 463 per 100 000 in the annually invited group and 368 per 100 000 in the control group who were invited in the third year only; the RR was 1·26 (95% CI 1·10–1·42).

All RRs in our sensitivity analyses were close to 1·00 (table 2), which suggests that no underlying incidence increase existed because of HRT exposure (sensitivity analysis C), unorganised screening (sensitiviy analyses B and C), migration, or other factors (sensitivity analyses A, B, and C; table 2).

To study possible time-related effects not related to screening, we used our staggered cohort method to analyse incidence data from the same period for women aged 40–69 years in four other counties (Värmland, Jämtland, Västerbotten, and Gotland) that had not yet started organised mammography screening. Using incidence data recorded in women aged 40–69 years at the start of the 6-year follow-up, the RR was 1·00 (95% CI 0·92–1·08; table 2).

To study whether an underlying incidence increase was present in the seven counties, we did two additional control tests. The first test was identical to that described above for women aged 40–47 years, but done for women aged 30–37 years. The other test was done with incidence data for women aged 40–69 years recorded 6 years earlier—ie, these data were collected before screening had begun. In both tests, neither study group nor control group were invited to screening. Therefore, a difference between the groups would indicate an underlying incidence change due to combined period and cohort effects. The RR for non-invited women aged 30–37 years in the seven counties included in this study was 1·01 (0·50–2·21). The RR for women aged 40–69 years at the start of the 6-year follow-up (and in the seven counties included in this study) was 1·01 (0·97–1·05) when data was collected in 1977–86, before the mammography screening programme started.

Discussion

Our findings show that the initiation of screening was associated with a substantial rise in the incidence of breast cancer in the screened group compared with the control group, which would be expected if we assume that screening allows earlier detection of breast cancers. However, after a round of prevalence screening in the control group there were still more cancers in the screened group, which suggests that some invasive breast cancers detected by repeated mammography

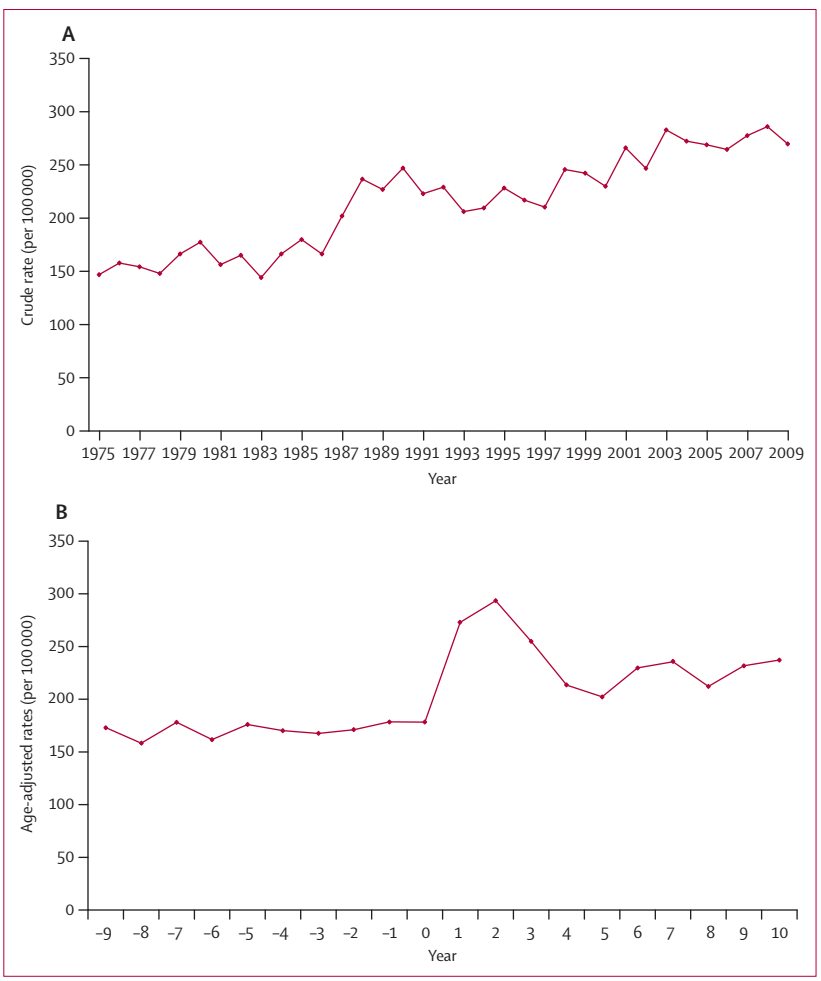

Figure 3: **Crude (A) and age-adjusted (B) breast cancer incidence rates for women aged 40–74 years**
For age-adjusted incidence rates (B) year 1 is the start of mammography screening.

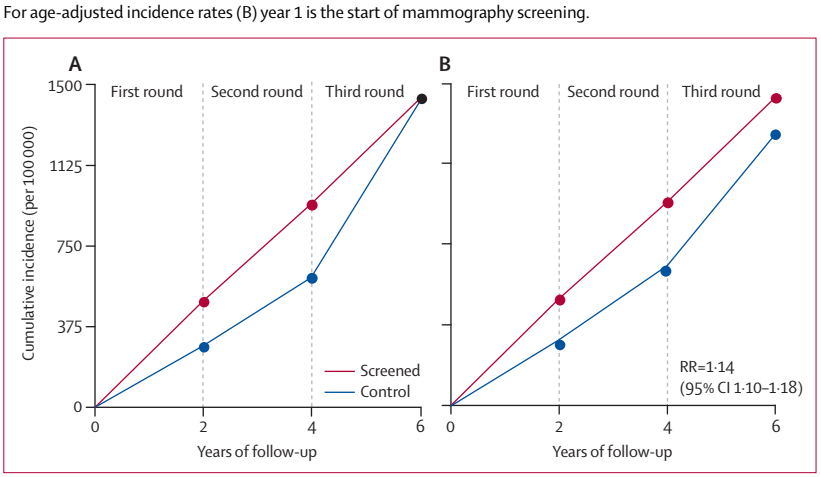

Figure 4: **Expected (A) and recorded (B) cumulative breast cancer incidence rates**
Incidences are given for the screened (red line) and control (blue line) groups.

	Control	Screening	Relative risk (screened *vs* control; 95% CI)
40–44 years*	753	912	1·21 (1·08–1·37)
45–49 years†	976	1117	1·14 (1·02–1·29)
50–54 years‡	1140	1352	1·19 (1·06–1·32)
55–59 years §	1315	1490	1·13 (1·02–1·25)
60–64 years ¶	1636	1848	1·13 (1·04–1·24)
65–69 years ||	1887	2186	1·16 (1·07–1·26)
40–69 years**	1269	1443	1·14 (1·10–1·18)

Finished screening at age: *45–49 years, †50–54 years, ‡55–59 years, §60–64 years, ¶65–69 years, ||71–75 years, and **45–75 years.

Table 1: **Effect of age (years) on the 6-year incidence of invasive breast cancer detection (cases per 100 000)**

	Years	Relative risk (95% CI)
A		
30–39 years	1977–86	1·11 (0·99–1·25)
B		
40–49 years	1977–86	1·04 (0·96–1·13)
50–59 years	1977–86	1·00 (0·93–1·07)
60–69 years	1977–86	1·01 (0·95–1·07)
C		
40–69 years	1981–90	1·00 (0·92–1·08)

(A) The results of a contemporary mock study done on the uninvited women aged 30–39 years in the seven counties (Uppsala, Jönköping, Kalmar, Örebro, Södermannland, Västernorrland, and Norrbotten) that provided data for our principal study in Sweden. (B) The results of a sensitivity study done on women aged 40–49 years, 50–59 years, 60–69 years in the seven counties but on data collected in 1977–86, before the mammography screening programme began. (C) The results of a mock study done with data collected between 1981 and 1990 for women aged 40–69 years in four counties (Gotland, Värmland, Jämtland, and Västerbotten) in Sweden with no screening programme.

Table 2: **Sensitivity analyses, by age (years)**

screening would not persist to be detected by screening at the end of 6 years. In other words, the natural course for some screen-detected breast cancers might be to spontaneously regress.

In our study, the screened group and the control group had very similar risk for the development of breast cancer because 26 of 30 cohorts in one group were also present in the other group. Moreover, although the oldest four cohorts in the screened group and the youngest four cohorts in the control group are unique to each group, the incidence rates before screening and the detection rates during screening in these cohorts were nearly identical to those seen in the most closely aged cohorts that were present in both groups, giving rise to similar RRs as recorded for cohorts present in both groups (figure 1 and table 1).

We previously proposed that most of the screening-related incidence increase is attributable to the diagnosis of invasive cancers that would have regressed spontaneously in the absence of screening.[2] A criticism of our study, other than the possible counfounding by use of HRT,[3] was that an analysis of Italian breast cancer incidence did not lend support to our suggestion.[9]

Compared with our previous analysis of Norwegian data,[2] our analysis of Swedish data had several strengths. First, the number of women assessed was three times larger. Second, the invited age group was much wider (40–74 years *vs* 50–69 years), which allowed us to extend the incidence data analysis to a mostly premenopausal age group (women younger than 50 years). Additionally, a larger proportion of women contributed information to both the screened and control groups, thereby reducing the possibility for confounding effects of fertility differences or other cohort-related factors. Third, the follow-up was much longer in this study than it was in our analysis of Norwegian data, allowing us to compare the incidence in the study group and the control group during the two screening rounds that followed at the end of the 6-year follow-up period. Fourth, our Swedish dataset is from a period with no underlying upward trend in breast cancer incidence—breast cancer incidence in Sweden was stable before the introduction of organised screening and it remained so during the entire study period in other Swedish counties that did not have organised screening. Fifth, the data allow analysis of the screening-related incidence change in a period when HRT use was rare. We have no specific data for the use of HRT in our cohorts, but according to Swedish drug statistics, the user frequency of HRT could have, at most, been 4% in the study group and 2% in the control group.[10] By contrast, in our analysis of Norwegian data,[2] the user frequency in the study group could have been 38%.[2] Even if all HRT users in the study group in this analysis were long-term users with 24% higher breast cancer risk (as in the Women's Health Initiative study[11]), and if HRT use in the control group had no effect on breast cancer incidence, this extreme assumption would explain only 1% of the recorded 14% higher cumulative cancer incidence in the study group.

Our results do not show that mammography screening is an imperfect technique—ie, that many breast cancers are missed at the first screening round but are detected at later screenings. Nor do they show an underestimation of the detection rate at the prevalence screening because of completion of screening later than we have assumed. In both cases, the absolute difference of 174 cancers in cumulative rates between the study group and the control group should be reduced with longer follow-up and additional screenings in both groups. However, when we included two extra rounds of screening in both groups (corresponding to 8 years and 10 years of follow-up), the absolute difference in cumulative rates was still very much the same—147 per 100 000 at 8 years and 181 per 100 000 at 10 years.

The null hypothesis that is refuted here is that the excess incidence of breast cancers in screened women is compensated for by a high detection rate when previously unscreened women undergo screening—ie, the null hypothesis states that the RR is 1·0 after screening of the

control group. The RR estimate is only indirectly associated to the extent of overdiagnosis. The estimated extent of overdiagnosis in the Swedish screening programme was reported to be about 35%,[1,5] and the data we present substantiate these findings (figure 3). However, because of our study design (with screening in both the test and control groups but with different intensity), the RR estimate was only 1·14. This finding shows that the RR estimate is sensitive to both the incidence rate at all screening rounds and the length of follow-up. Low detection rates at the prevalence screening (as seen in this study for the women aged 40–49 years) increase the overall RR, whereas high detection rates at the prevalence screening (as seen in this study for aged 50–74 years) decrease the overall RR. High detection rates at later screening rounds also increase the overall RR estimate. The RR would also have increased if we had extended follow-up and included more screening rounds before the control group was invited to a prevalence screen; it would have tended to 1·35. Therefore, our findings suggest that the natural history of most of the 35% extra cancers detected in the women invited to screening is to undergo spontaneous regression, as also shown by the detection rate at the prevalence screen of women aged 73–74 years. At this screening, only a very small proportion of the excess of cancer from age 40–74 years in screened women seem to be present in previously unscreened women aged 73–74 years (figure 2).

Our interpretation of the incidence data is also supported by the independent finding that the decrease in the incidence of clinical cancers in women that are no longer invited to screening (women aged 75–79 years) is too small to compensate for more than a little of the long-term increase during the years the women were invited to screening.[1,2]

We recorded no underlying incidence increase in our Swedish dataset, as shown in our sensitivity analyses of counties where screening had not yet started. However, a 0·8% incidence increase of breast cancer not caused by the organised screening programme was calculated by others for Sweden in 1972–85.[12] This increase might be partly explained by opportunistic screening in urban regions and by the five Swedish randomised screening trials that included 170 000 women (12% of the female population aged 40–74 years) from 1977 to 1991. However, if we assume that a 0·7% underlying incidence increase existed in our dataset (figure 3), then our RR would be 1·11.

In our analyses of the dataset from Norway,[2] we were concerned that opportunistic screening in the control group could have affected our results. By contrast, in the present analysis of the Swedish dataset, the incidence of breast cancer was very stable in women aged 40–74 years in the 10 years before screening started (figure 3).

A potential confounder that we cannot completely rule out is the possibility that screening sensitivity increases with time. However, if this were an important confounder we would expect to see pronounced time trends towards increasing detection rates at the prevalence screen, decreasing diameters of tumours detected at any screening round, decreasing incidence and diameters of interval cancers, and decreasing incidence in women older than 74 years. But such changes have been only minor.[5] Indeed, in our dataset, the incidence rate at the prevalence screening of women aged 39–41 years was constant during the ten screening rounds (93 per 100 000).

Although we cannot completely rule out that, in combination, small effects of several of the factors discussed above (ie, increasing HRT use with time, other temporal and cohort effects, and increasing screening sensitivity with time) might give rise to higher breast cancer incidence in the screened group than in the control group, we think that the effect of such confounding is probably not large and that this confounding would not be age and cohort independent (table 1).

In Norway, screening is associated with a substantial increase in the proportion of early (stage 1) breast cancer, but no proportional reduction occurs in the absolute number of higher stage breast cancers.[12] Screening in Sweden will probably have a similar effect, but data from the Swedish Cancer Registry does not include information on tumour stage.

Analyses similar to ours can be done on datasets from any screening programmes with high levels of overdiagnosis. For example, Fryback and colleagues[13] used a stochastic simulation model to replicate breast cancer incidence and mortality rates in the USA between 1975 and 2000, including the 1980s when mammography screening spread. To fit observed statistics,[13] it was

***Panel:* Research in context**

Systematic review

We searched PubMed with the terms "spontaneous regression" and "breast cancer" and got 208 results (date of last search Aug 24, 2011). We found one epidemiological study testing the hypothesis that some invasive breast cancers spontaneously regress[2] and one simulation study of breast cancer growth rates estimating the frequency of regression.[13] Analysis of data from the follow-up of the randomised mammography screening trials in Malmö, Sweden,[14] and in Canada[16,17] shows that some of the extra cancers detected in women invited to screening were not compensated for when women in the control group were screened at a later time. This suggests that some invasive breast cancers detected in the screening groups would never have presented in the control arms. A cohort study[2] of the cancer incidence rates in Norway during the introduction of mammography screening suggests that the natural history of some screen-detected invasive tumours would be to undergo spontaneous regression if they had been left untreated.

Interpretation

In our study, cumulative cancer incidence during a 6-year period was significantly higher in the study group invited to regular screening than it was in the control group that had one prevalence screen at the end of the 6-year period. On the basis of our findings, we believe that, if left untreated, the natural history of many screen-detected invasive cancers is to undergo spontaneous regression. Detection of cancers that would have otherwise undergone spontaneous regression could explain almost all the increase in incidence noted when mammography screening is done.

necessary to postulate that about 40% of initiated breast cancer were of limited malignant potential—ie, tumours that "progress to a maximum of approximately 1 cm diameter, dwell at this size for 2 years, and then regress if undetected". These calculations also rely on the fact that tumours that are detectable by mammography screening do not accumulate in the absence of screening, but regress and are undiagnosed.

The randomised trials of screening were not designed to study the possibility of spontaneous regression of screen-detected cancers. But the Malmö trial[14] provides useful data. It is one of the most reliable trials[15]—it ran for 9 years and, at 15 years, has the longest follow-up of all the trials after the randomised phase ended (panel). The age group 45–54 years is of interest, because women in both the study and the control group were included in the service screening programme after the trial ended, from 1990 to 2001. The RR of the cumulative incidence of invasive cancer for women aged 45–54 years was 1·16 (95% CI 0·98–1·36) at the end of the trial period.[14] The extra cancers detected in the randomised phase of the trial were not compensated for later, when women in the control group enter the national mammography screening programme. The fact that extra cancers were not compensated for shows that a substantial proportion of invasive breast cancers detected in the screening group would not have been detected in the control group, and also that some of the screening-detected breast cancers would have regressed spontaneously. Two Canadian trials[16,17] reported that a proportion of invasive breast cancers detected in the screening group would not have been detected in the control group, despite screening for 4 years after the end of the trial. In these trials, a truly unscreened control group was available only for women aged 40–49 years and for this age range the excess incidence in the screened group was 22% (RR 1·22, 95% CI 1·09–1·37), which accords with our findings for this age group.

Welch[18] has argued that the amount of overdiagnosis depends on a mammographer's threshold at which they recommend a biopsy. He advised that randomised controlled trials should be done to define higher thresholds for the size of breast masses for which biopsy should be recommended. We support his proposition. Such a trial can also be designed to confirm, by direct observation, that some tumours are not detected in mammograms subsequent to the one in which they are first detected. Our findings, as well as Fryback and colleagues',[13] suggest that few patients are needed to obtain direct observation of spontaneous regression during a 1–2 year follow-up period. We propose that after needle biopsy of a small (<10 mm) oestrogen-receptor-positive cancer detected in screening, a patient should be invited to be treated with tamoxifen or aromatase inhibitors during continuous MRI or ultrasound monitoring of the size of the lesion. If tumour progression occurs, a patient should undergo immediate surgery. If no growth or tumour regression occurs surgery could be postponed. Another possible study is the sampling of tumour tissue and blood at optimum times to study local or systemic (including hormonal and immunological) mechanisms that might be involved in tumour growth reversal.

Contributors

P-HZ collected the dataset, wrote the first draft, and did the statistical analyses. All authors contributed to the design and writing of the paper.

Conflicts of interest

We declare that we have no conflicts of interest.

References

1 Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009; **339:** b2587.

2 Zahl P-H, Mæhlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 2008; **168:** 2311–16.

3 Weedon-Fekjær H, Ragnhild Sørum R, Brenn MK. Hormone therapy use may explain recent results regarding tumor regression. *Arch Intern Med* 2009; **169:** 996–97.

4 Barlow L, Westergren K, Holmberg L, Talbäck M. The completeness of the Swedish Cancer Register – a sample survey for year 1998. *Acta Oncologica* 2009; **48:** 27–33.

5 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:** 842–47.

6 Olsson S, Andersson I, Karlberg I, Bjurstam N, Frodis E, Håkansson S. Implementation of service screening with mammography in Sweden: from pilot study to nationwide programme. *J Med Screen* 2000; **7:** 14–18.

7 Statistics Sweden. http://www.ssd.scb.se/databaser/makro/start.asp (accessed Aug 24, 2011).

8 Efron B, Tibshirani RJ. An introduction to the Bootstrap. NY: Chapman and Hall, 1993.

9 Paci E, Puliti D, Zappa M. Mammographic screening and breast cancer: florentine data. *Arch Intern Med* 2009; **169:** 997–98.

10 Lambe M, Wigertz A, Holmqvist M, et al. Reduction in use of hormone replacement therapy: effects on Swedish breast cancer incidence trends only seen after several years. *Breast Cancer Res Treat* 2010; **121:** 679–83.

11 Chlebowski RT, Hendrix SL, Langer RD, et al. Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative randomized trial. *JAMA* 2003; **289:** 3243–53.

12 Zahl P-H, Strand BH, Mæhlen J. Breast cancer incidence in Norway and Sweden during introduction of nation-wide screening: prospective cohort study. *BMJ* 2004; **328:** 921–24.

13 Fryback DG, Stout NK, Rosenberg MA, et al. The Wisconsin breast cancer epidemiology simulation model. *J Natl Cancer Inst Monogr* 2006; **36:** 37–47.

14 Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of overdiagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* 2006; **332:** 689–91.

15 Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2009; **4:** CD00287.

16 Miller AB, To T, Baines CJ, Wall C. National breast screening study-2: 13-year results of a randomized trial in women aged 50-59 years. *J Natl Cancer Inst* 2000; **92:** 1490–99.

17 Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up: a randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* 2002; **137:** 305–12.

18 Welch HG. Overdiagnosis and mammography screening. *BMJ* 2009; **339:** b1425.