Overdiagnosis due to breast cancer screening: updated estimates of the Helsinki service study in Finland

S Heinävaara*,[1], T Sarkeala[1] and A Anttila[1]

[1]Finnish Cancer Registry, Mass Screening Registry, Unioninkatu 22, 00130 Helsinki, Finland

Background: Overdiagnosis is the most important adverse event of breast cancer screening with the estimates ranging from 0% to 40–50% depending on invitational age and methods. We updated the estimates of overdiagnosis in Helsinki service screening study in Finland by comparing the observed and expected cumulative incidence of all breast carcinomas and invasive breast carcinomas.

Methods: Women aged 50–59 years have been invited to Helsinki service screening since 1986. The incidence of breast carcinoma in the first invited birth cohorts born in 1935–1939 was compared with older, non-invited cohorts. The minimum follow-up time of the invitees after the last screening round was 14 years. Expected cumulative incidence rates were estimated with two alternative approaches.

Results: For both any breast carcinoma and invasive breast carcinoma, the estimates of overdiagnosis varied from 5% (95% CI = −1, 11%) to 7% (95% CI = 1, 13%) depending on the approach.

Conclusions: Our estimates of overdiagnosis are of the same magnitude than other plausible estimates in Europe. Both alternative approaches produced similar estimates for the expected cumulative incidence, which increased the confidence in the estimates of overdiagnosis.

Overdiagnosis is the most important adverse effect of breast cancer screening (International Agency for Research on Cancer, 2002). The concept includes (1) unnecessary diagnosis and treatment of breast cancers that are not destined to cause a death or symptoms, to (2) diagnosis of breast cancer at screening, histologically confirmed, that would never have been identified clinically in the lifetime of the woman (International Agency for Research on Cancer, 2002; Welch and Black, 2010; Puliti et al, 2012a). Following the first definition, overdiagnosis can occur in screening or outside it, but a reliable quantification of its size is not possible. For the second definition, the overdiagnosis can be quantified by comparing observed and expected cumulative incidence of breast cancer several years after screening (Puliti et al, 2012a).

Overdiagnosis can ideally be estimated from randomised screening trials where the observed breast cancer incidence can be derived from the intervention arm and the expected incidence without screening from the control arm. However, data from randomised clinical trials are limited to three studies only (Miller et al, 2000, 2002; Zackrisson et al, 2006), and therefore most estimates for overdiagnosis are based on service screening studies. In the service screening setting, the observed breast cancer incidence can be drawn from the population subjected to screening, or actually screened, but the expected incidence needs to be estimated. This estimation is based either on non-invitees who are comparable to invitees with respect to age, calendar time and area as closely as possible, or if the entire target population is invited to screening, by extrapolating incidence trends from pre-screening era. Nevertheless, assessment of temporal trends is difficult if the whole target population has been invited, as changes in breast cancer incidence are likely to be dependent on changes in breast cancer diagnostics and risk factors.

As the scope of screening is to lower mortality from breast cancer by introducing early diagnosis, breast cancer incidence will inevitably increase shortly after the onset of screening. In later years, that is, during the intervals between the screens and after the last screen, the incidence can be expected to be lower than without screening (International Agency for Research on Cancer, 2002; Puliti *et al*, 2012a). The changes in breast cancer incidence because of lead time should be taken into account when estimating the expected incidence without screening (Puliti *et al*, 2012a). Alternative methods have been suggested for lead time adjustment (Duffy *et al*, 2008; Gøtzsche *et al*, 2009; Duffy and Parmar, 2013) as well as a concern on overadjustment (Zahl *et al*, 2013).

In previous European studies, a credible overdiagnosis estimate of 1–10% has been suggested (Duffy *et al*, 2010; de Gelder *et al*, 2011; Puliti *et al,* 2012b; Njor *et al*, 2013). Without lead time correction, the differences between the estimates would have been wider (Puliti *et al*, 2012a). As overdiagnosis may increase with age (Biesheuvel *et al*, 2007), estimates can differ because of various age groups invited to screening. Moreover, as there is no agreement how to measure overdiagnosis, differences between estimates can also be due to different methods. Therefore, estimates with a wide range have been presented (Marmot *et al*, 2013) as well as some dispute about the magnitude of the problem (Gøtzsche *et al*, 2012).

The earlier Finnish study from Helsinki area reported 18% difference between the observed and expected cumulative incidence few years after the last screen in women invited biennially to service screening in ages 50–59 years (Anttila *et al*, 2002). The Helsinki screening programme provides a unique natural experiment for the estimation of overdiagnosis as women aged 50–59 years were gradually invited to the screening programme, and the target age group of the programme remained unchanged until the end of 2006.

The aim of the current study is to update the overdiagnosis estimates of the Helsinki study with a longer follow-up. Expected incidence of breast cancer without screening is modelled with two alternative approaches and the alternatives are compared with each other. In the primary analysis, the overdiagnosis is evaluated for all breast carcinomas and invasive breast carcinomas. In the secondary analysis, the overdiagnosis is estimated by stage of invasive breast carcinoma.

MATERIALS AND METHODS

The Helsinki service screening programme was targeted to 50- to 59-year-old women, who were invited to screening every 2 years. The programme started in 1986 by inviting one birth cohort (born in 1936) and expanded gradually to include the whole target age (Figure 1). The women born in 1935–1939 were the first invited 5-year birth cohort and were invited to screening in Helsinki in 1986–1997. The women born in 1934 or earlier did not receive invitation in their lifetime.

To analyse the incidence of breast cancer, data of women resident in Helsinki and diagnosed with breast carcinoma in 1970–2011 were obtained from the Finnish Cancer Registry (FCR). The corresponding annual mean population counts by age and calendar year were received from the Helsinki Registry Office (for 1970–1975) and from the Statistics Finland (for 1976–2011). For the primary analyses, the compiled data set included all breast carcinomas, invasive breast carcinomas and mean population counts tabulated by 5-year birth cohorts (1920–1924, 1925–1929, 1930–1934, 1935–1939), 5-year age groups (40–44, 45–49, …, 70–74). All breast carcinomas included also ductal carcinomas *in situ* (∼88% of *in situ* carcinomas) and lobular carcinomas *in situ*. For the secondary analyses, the compiled data set included invasive breast carcinomas by stage (localised, non-localised, unknown) and
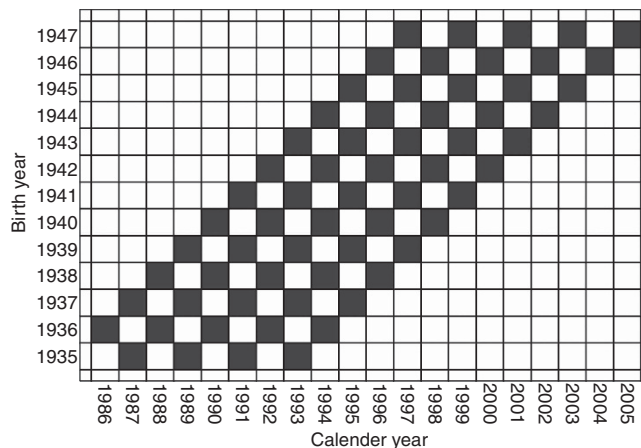

Figure 1. The design of the Helsinki service screening for first invited birth cohorts (1935–1947) from 1986 to 2005. The black box indicates the invited women in a given calendar year and age, and the white box the corresponding uninvited women.

mean population counts. The non-localised stage included carcinomas that have metastasised to regional lymph nodes or further, grew into neighbouring tissue or were known to be spread but not known how far; the localised stage included localised carcinomas and the unknown stage includes the rest. The classification of stage used in the FCR cannot be converted into TNM stage.

The data were constructed symmetrically for each 5-year birth cohort. All annual cohorts were followed until the age of 72 years, the oldest four of five birth cohorts were followed until the age of 73 years and the oldest three until the age of 74 years. For the cohort born in 1935–1939, the minimum follow-up time after the last screening round at the age of 58 years in 1997 (Figure 1) was thus 14 years.

Incidence of breast carcinoma was modelled with the Poisson regression using 5-year age group, 5-year birth cohort and a 'screening' as categorical variables. The 'screening' variable was used for the lead time adjustment, that is, by allowing an increase in the incidence in ages 50–59 years and a decrease in ages 60–64 years in the 5-year birth cohort invited to screening (1935–1939). Thus, assuming that the effect of age on the incidence does not differ between the birth cohorts, the 'screening' variable should capture changes in the incidence because of invitation to screening in the age groups of 50–59 and 60–64 years in the birth cohort of 1935–1939. Further, the 'screening' variable is describing the effect of screening on the incidence if there were no other changes occurring among 50- to 64-year olds from 1986 onwards. The nearest three non-invited 5-year birth cohorts (1920–1924, 1925–1929, 1930–1934) were included in the model to stabilise the estimation of age effects. The further details with the formula of the model are shown in Appendix 1.

The potential overdiagnosis was studied by comparing the observed and expected cumulative incidences. We estimated the observed cumulative incidence by summing the incidence in the invited 5-year birth cohort (1935–1939) from age 50 years up to age 74 years. The expected incidence without screening was estimated using two alternative approaches. First, it was calculated for the nearest, non-invited 5-year birth cohort (1930–1934) by correcting its observed incidence with the model-based birth cohort effect. In other words, the observed incidence in the 5-year birth cohort 1930–1934 was corrected by the difference in the incidence between the 5-year cohorts 1930–1934 and 1935–1939. This approach is denoted by A1. Second, the expected incidence was calculated for the invited 5-year cohort born in 1935–1939 by correcting the observed incidence by the model-based 'screening'

effect in ages 50–59 and 60–64 years. This is to say that we removed the effect of 'screening' in ages 50–59 and 60–64 years from the observed incidence in the cohort of 1935–1939. This latter approach is denoted by A2. The formulae for calculating expected incidence rates A1 and A2 are given in Appendix 1.

The estimates of overdiagnosis (in % with 95% confidence intervals (CI)) were derived by modelling the ratio of observed and expected numbers of breast carcinoma using the Poisson regression. Please note that the ratio of observed and expected cumulative incidence rates is not exactly the same as that of observed and expected numbers.

The primary analyses were performed for all breast carcinomas and invasive breast carcinomas, and secondary analysis for localised and non-localised invasive breast carcinomas. Invasive breast carcinomas with unknown stage were relatively rare and were therefore excluded from the stage-specific analyses.

All the analyses were performed with Stata version 12 (StataCorp., 2011).

RESULTS

The data included almost 1 000 000 women years (Table 1). The frequencies of any breast carcinoma and invasive breast carcinoma peak in the invited cohort (1935–1939) in the age group 50–54 years and decreases thereafter in the age group 60–64 years.

Incidence rate ratios (IRRs) of 'screening' in the age group 50–59 and 60–64 years were the same for any breast carcinoma and invasive breast carcinoma. In the age group of 50–59 and 60–64 years, the IRRs were 1.25 (95% confidence interval (CI) = 1.09, 1.44) and 0.86 (95% CI = 0.86, 1.03), respectively, within the invited birth cohort. (These IRRs are used in Appendix 1 in the alternative A2.) For a change in the incidence between the cohorts of 1930–1934 and 1935–1939, IRRs of the invited cohort were 1.09 (95% CI = 0.99, 1.20) and 1.07 (95% CI = 0.97, 1.18) for any breast carcinoma and invasive breast carcinoma, respectively, compared with the last non-invited birth cohort (1930–1934). (These IRRs equal to coh^{est} in Appendix 1 used in the alternative A1.)

The observed and expected cumulative incidence rates for 50- to 74-year-old women are illustrated for any breast carcinoma in Figure 2. The expected cumulative incidence rates A1 and A2 are generally close to each other but expected cumulative incidence A1 remains slightly lower than A2 in the oldest ages. The observed and expected cumulative incidence rates A1 and A2 were 9.73/1000, 9.22/1000 and 9.34/1000, respectively. The estimate of over-diagnosis was 7% (95% CI = 1, 13%) for the alternative A1 and 5% (95% CI = − 1, 11%) for the alternative A2. Observed and

expected cumulative incidence rates A1 and A2 are lower for invasive breast carcinoma (9.17/1000, 8.64/1000 and 8.81/1000, respectively, and Figure 3), but the estimates of overdiagnosis are the same than those for any breast carcinoma.

In the secondary analyses by the stage of invasive breast carcinoma, the expected increase and decline in the incidence because of screening in the age groups of 50–59 and 60–64 years are seen in the localised stage (Table 2). The frequencies of unknown stage remain at a relatively stable level in all the age groups and in the both cohorts (overall figure being 4% for the both cohorts).

For cases confined to the localised invasive breast carcinoma, IRRs of 'screening' were 1.63 (95% CI = 1.36, 1.96) and 0.81 (95% CI = 0.63, 1.02) in the age groups 50–59 and 60–64 years, respectively, within the invited birth cohort. For cases with the non-localised invasive breast carcinoma, the corresponding IRRs were 0.84 (95% CI = 0.65, 1.07) and 1.00 (95% CI = 0.75, 1.32), respectively. Incidence rate ratios of the invited cohort were 1.10 (95% CI = 0.96, 1.25) and 1.01 (95% CI = 0.86, 1.17), respectively, for the localised and non-localised invasive breast carcinoma compared with the last non-invited birth cohort (1930–1934).

The overdiagnosis of cases confined to the localised invasive breast carcinoma was 15% (95% CI = 7, 24%) for the both alternatives. The estimates of overdiagnosis of cases confined to the non-localised invasive breast carcinoma were 0% (95% CI = − 9, 10%) and − 5% (95% CI = − 14, 4%), respectively, for the alternatives A1 and A2. The observed and expected cumulative incidence rates for women with the localised or non-localised invasive breast carcinoma are illustrated in Supplementary Material.
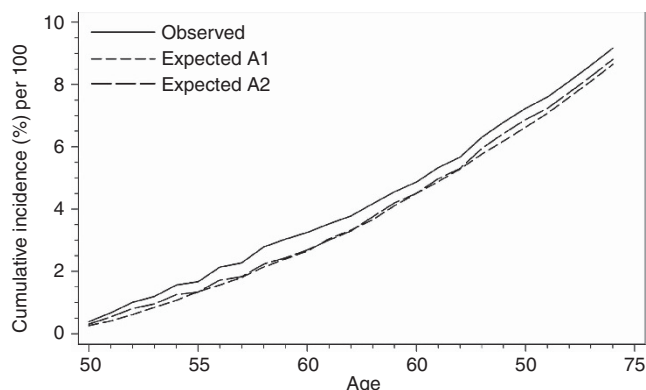

Figure 2. Observed and expected cumulative incidence of any breast carcinoma for women aged 50–74 years. Expected cumulative incidence was estimated with two alternative approaches (A1 and A2).

Table 1. Frequencies (and percentages) of all breast carcinomas, invasive breast carcinomas and mean population counts for the last non-invited birth cohort of 1930–1934 and for the first invited birth cohort of 1935–1939 by 5-year age group

Age group (years)	Any breast carcinoma		Invasive breast carcinoma		Mean population	
	1930–1934	1935–1939	1930–1934	1935–1939	1930–1934	1935–1939
40–44	77 (6)	68 (5)	77 (6)	66 (5)	77 418 (16)	79 195 (16)
45–49	122 (10)	137 (9)	122 (10)	133 (10)	73 739 (16)	77 429 (16)
50–54	147 (12)	246 (17)	144 (12)	236 (17)	72 212 (15)	75 627 (15)
55–59	173 (14)	228 (15)	172 (15)	216 (15)	69 123 (15)	73 308 (15)
60–64	213 (17)	223 (15)	207 (17)	212 (15)	64 910 (15)	69 945 (14)
65–69	250 (20)	316 (21)	237 (20)	289 (21)	60 944 (13)	64 914 (13)
70–74	241 (20)	259 (18)	226 (19)	246 (18)	49 850 (11)	52 825 (11)
Sum	1223 (100)	1477 (100)	1185 (100)	1398 (100)	468 193 (100)	493 242 (100)

DISCUSSION

We estimated the overdiagnosis of breast carcinoma among 50- to 59-year old women to be around 5–7% after the adjustment of underlying breast cancer risk and lead time. Our estimates are of the same magnitude than other plausible estimates in other organised screening studies in Europe (Duffy *et al,* 2010; de Gelder *et al,* 2011; Puliti *et al,* 2012b; Njor *et al,* 2013). Our data covered 25 years (1986–2011) with biennial mammography screening since the start of the routine screening service in Helsinki. The cohort born in 1935 received the last invitation in 1993 and that born in 1939 in 1997. The study thus covers the minimum of 14 years of follow-up after the last invitation round. Our study period should therefore be long enough to take adequately adjusted for the lead time bias, that is, a minimum of two decades and at least 10 years after the last screen (Duffy and Parmar, 2013).

The Helsinki service study offered a unique opportunity to study the overdiagnosis due to breast cancer screening among 50- to 59-year-old women. However, as the overdiagnosis may increase with age (Biesheuvel *et al,* 2007), our estimates could have been higher, if older age groups had been included. Unfortunately, invitational age groups have been changing with time in other municipalities and therefore the estimation of overdiagnosis for a wider age group (say 50–69 years) is not possible without unverifiable assumptions. As women living in Helsinki have easy opportunities to opportunistic screening and knowing that their acquaintances living in other municipalities are being invited also

Figure 3. Observed and expected cumulative incidence of invasive breast carcinoma for women aged 50–74 years. Expected cumulative incidence was estimated with two alternative approaches (A1 and A2).

after their 60th birthday, they may well have done so. Population-based screening programme itself may have also increased the attendance to opportunistic screening (Boncz *et al,* 2008). As the attendance to opportunistic mammography screening is not registered in Finland and its magnitude is thus unknown, it could not be taken into account in the analyses. However, the percentage of *in situ* carcinomas of all breast carcinomas is an indicator of opportunistic screening. In the age group 60–64 years, the percentage of *in situ* carcinomas in the birth cohorts of 1930–1934 and 1935–1939 were 2.8% and 4.9%, respectively (from Table 1). In the age group 65–69 years, the corresponding percentages were 5.7% and 7.0%, respectively. It therefore seems that previously invited women may have continued to be screened after the organised screening programme on their own cost. This could at least partly explain an unexpected increase in the incidence of all breast carcinomas and invasive breast carcinomas among 65- to 69-year olds in the birth cohort of 1935–1939 (Table 1). If so, our estimate of overdiagnosis can be an overestimate for 50- to 59-year-old women. On the other hand, in the age group 50–59 years, the percentages of *in situ* carcinomas in the birth cohorts 1930–1934 and 1935–1939 were 1.3% and 4.6%, respectively. Therefore it is quite possible that also the non-invited birth cohorts could have attended to opportunistic screening in the older age groups (60–69 years). Overall, as long as the attendance to opportunistic screening is not known, the effect of opportunistic screening on our estimates of overdiagnosis remains unclear.

In Finland, the percentage of screen-detected *in situ* carcinomas among 50- to 64-year-old women in 1991–2000 was 10% (4–18%) and is in line with the desirable level (10–20%; Sarkeala *et al*, 2004). This figure was based on the first diagnosis after referral examination registered at the Mass Screening Registry. In the current study, data of breast carcinoma were from the FCR. Finnish Cancer Registry collects information from various sources and has complete follow-up of incident breast cancer cases. Therefore, diagnoses may change with time. Also, FCR has practise of coding a screen-detected carcinoma *in situ* as invasive (with original data of diagnosis), if the lesion progresses invasive with a 2-year period (Sarkeala *et al*, 2006). In earlier studies, the percentage of screen-detected *in situ* carcinomas among 50- to 64-year-old women was seen to decrease from 10% to about 5% at FCR (Sarkeala *et al*, 2006). This explains the low percentage of *in situ* carcinomas in our data. However, these reasons are unlikely to have an effect on our estimate of overdiagnosis.

Individual data of screening invitations and participations were not available for the first years of screening programme in Helsinki. As the participation to mammographic screening has been high
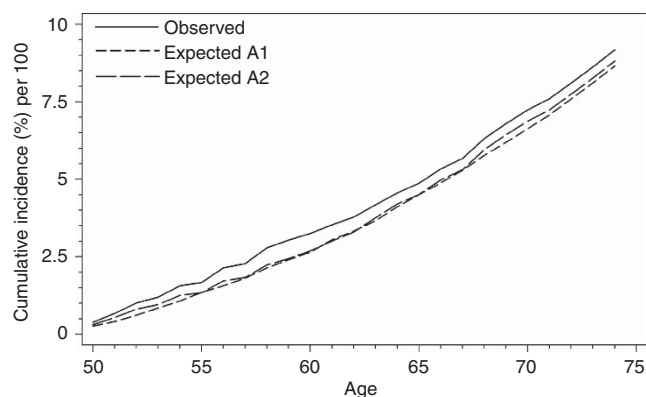
	Localised		Non-localised		Unknown	
Age group (years)	**1930–1934**	**1935–1939**	**1930–1934**	**1935–1939**	**1930–1934**	**1935–1939**
40–44	32 (5)	36 (4)	37 (7)	27 (5)	8 (15)	3 (5)
45–49	65 (10)	74 (9)	53 (11)	45 (9)	4 (8)	14 (25)
50–54	80 (13)	171 (20)	52 (10)	60 (12)	12 (23)	5 (9)
55–59	100 (16)	154 (18)	68 (14)	58 (11)	4 (8)	4 (7)
60–64	135 (21)	119 (14)	69 (14)	86 (17)	3 (6)	7 (13)
65–69	118 (19)	160 (19)	108 (22)	111 (22)	11 (21)	18 (33)
70–74	107 (17)	124 (15)	109 (22)	118 (23)	10 (19)	4 (7)
Sum	637 (100)	838 (100)	496 (100)	505 (100)	52 (100)	55 (100)

Table 2. Frequencies (and percentages) of invasive breast carcinoma by stage for the last non-invited birth cohort of 1930–1934 and first invited birth cohort of 1935–1939 by 5-year age group

Table 3. The frequencies (and percentages) of any breast carcinoma, invasive breast carcinoma and *in situ* breast carcinoma in four age groups for the last non-invited birth cohort of 1930–1934 and the first invited birth cohort of 1935–1939

Age group (years)	Any breast carcinoma		Invasive breast carcinoma		*In situ* carcinoma	
	1930–1934	1935–1939	1930–1934	1935–1939	1930–1934	1935–1939
40–49	199 (16)	205 (14)	199 (17)	199 (14)	0 (0)	6 (8)
50–59	320 (26)	474 (32)	316 (27)	452 (32)	4 (11)	22 (28)
60–64	213 (17)	223 (15)	207 (17)	212 (15)	6 (16)	11 (14)
65–74	491 (40)	575 (39)	463 (39)	535 (38)	28 (74)	40 (51)
All	1223 (100)	1477 (100)	1185 (100)	1398 (100)	38 (100)	79 (100)

(82% in 1986–1997) in Helsinki (Anttila *et al*, 2002), invitation and participation to screening are closely related. However, some uncertainty will remain whether the increase in the incidence of any breast carcinoma and invasive breast carcinoma in the age group 50–59 years and in the cohort of 1935–1939 is really due to the organised screening. There may have been a change in risk factors or early diagnostics. For example, the use of hormone replacement therapy increased more in Helsinki than in other areas of Finland in the late 1980s and seemed to differ between age groups (Topo *et al*, 1991). Early diagnostics is likely to affect the incidence of localised rather than non-localised invasive breast carcinoma. The difference in the incidence between the cohorts of 1930–1934 and 1935–1939 was estimated to be around 10% for the localised invasive breast carcinoma and 1% for the non-localised invasive breast carcinoma. The incidence of breast carcinoma in the age group 50–59 years in the invited 5-year cohort is thus likely to be affected both by the organised screening programme and early diagnostics. However, if early diagnostics is not dependent on age, it should not affect our estimates of overdiagnosis.

Our results are in concordance with the earlier reported results (18%; Anttila *et al*, 2002) for the first invited birth cohort born in 1935–1939 and living in Helsinki at the time of diagnosis. In that study, the end of the follow-up was in 1997 indicating that the only the oldest cohort born in 1935 achieved the age of 62 years. Corresponding estimates of overdiagnosis calculated until the age of 62 years were 14% (95% CI = 4, 24%) for alternative A1 and 15% (95% CI = 6, 25%) for alternative A2.

It was expected *a priori* that screening increases the breast cancer incidence during the screens and decreases after the last screen (International Agency for Research on Cancer, 2002; Puliti *et al*, 2012a). The increase was assumed to occur in the ages 50–59 years and the decrease in the age group 60–64 years. The compensatory drop was restricted to age group 60–64 years as the mean sojourn time has been estimated to be about 2 years (Wu *et al*, 2010). Additional explorations also showed that the further 5-year age group after the last screen (ages 65–69 years) or 10-year age group after the last screen (ages 60–69 years) would not have had an effect on the incidence (results not shown). As the data were aggregated by 5-year age groups and birth cohorts, decreases between the biannual screens could not be taken into account. The expected decline, compensatory drop, in the age group 60–64 years was marginally non-significant for any breast carcinoma and invasive breast carcinoma. However, our results are in line with earlier reported study, which was based on data of about 60% of Finnish municipalities excluding Helsinki (Seppänen *et al*, 2006).

We used two alternative approaches for estimating the expected incidence without screening (A1 and A2). If the model had not fitted well the incidence data, we might have seen a visible difference between the alternatives A1 and A2 even if cumulative incidence is a robust measure levelling off random fluctuations in annual incidence of breast carcinoma. Overall, the differences between the expected cumulative incidence rates A1 and A2 by the

age of 74 years were small indicating that the results are stable and not dependent on the chosen basis for the estimation of the expected incidence, that is, the observed incidence in the last non-invited 5-year cohort or in the invited 5-year cohort. We can thus be quite confident with the extrapolation, within a geographical area, with respect to age and cohort effects.

The estimates of ovediagnosis were the same for any breast carcinoma and invasive breast carcinoma. As we had sufficiently long follow-up time after the last screen, the differences between the cumulative incidence of any breast carcinoma and invasive breast carcinoma levelled off with age. This levelling off hides the changes in incidence because of an early diagnosis in a pre-invasive phase in an invited birth cohort, that is, an expected increase *in situ* carcinoma and a decline in invasive breast carcinoma after a last invitation. We can see the increase in the incidence of *in situ* carcinoma in the invited age group but not decline (Table 3). The latter is not surprising as the percentages of *in situ* carcinomas of all breast carcinomas were small (3–7%).

To conclude, our estimates of overdiagnosis due to breast cancer screening among 50- to 59-year-old women were less than 10%. Even if our target age group would have wider and older, and our estimates would have been higher, the overdiagnosis of 25–30% would have been out of the credible range.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Anttila A, Koskela J, Hakama M (2002) Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. *J Med Screen* **9**: 153–158.

Biesheuvel C, Barrett A, Howard K, Houssami N, Irwig L (2007) Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol* **8**: 1129–1138.

Boncz I, Sebestyén A, Pintér I, Battyány I, Ember I (2008) The effect of an organized, nationwide breast cancer screening programme on non-organized mammography activities. *J Med Screen* **15**: 14–17.

de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ (2011) Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* **33**: 111–121.

Duffy SW, Lynge E, Johsson H, Ayyaz S, Olsen AH (2008) Complexities in the estimation of overdiagnosis in breast cancer screening. *Br J Cancer* **99**: 1176–1178.

Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen TH, Yen AM, Smith RA (2010) Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* **17**: 25–30.

Duffy SW, Parmar D (2013) Overdiagnosis in breast cancer screening: the importance of length of observation and lead time. *Breast Cancer Res* **15**: R41.

Gøtzsche PC, Jørgensen KJ, Zahl P-H, Mæhlen J (2012) Why mammography screening has not lived up to expectations from the randomised trials. *Cancer Causes Control* **23**: 15–21.

Gøtzsche PC, Jørgensen KJ, Mæhlen J, Zahl P-H (2009) Estimation of lead time and overdiagnosis in breast cancer screening. *Br J Cancer* **100**: 219.

International Agency for Research on Cancer (2002) *Breast Cancer screening IARC Handbooks of Cancer Prevention* vol 7IARC: Lyon.

Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M.. The Independent UK Panel on Breast Cancer Screening (2013) The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* **108**: 2205–2240.

Miller AB, To T, Baines CJ, Wall C (2002) The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. *Ann Intern Med* **137**: E305–E308.

Miller AB, To T, Baines CJ, Wall C (2000) Canadian National Breast Screening Study-2: 13-year results of a randomised trial in women aged 50-59 years. *J Natl Cancer Inst* **92**: 1490–1499.

Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E (2013) Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ* **346**: f1064.

Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, Paci E. EUROSCREEN Working Group (2012a) Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* **19**: 42–56.

Puliti D, Miccineci G, Zappa M, Mannesci G, Crocetti E, Paci E (2012b) Balancing harms and benefits of service mammography screening programs: a cohort study. *Breast Cancer Res* **14**: R9.

Seppänen J, Heinävaara S, Anttila A, Sarkeala T, Virkkunen H, Hakulinen T (2006) Effects of different phases of an invitational screening program on breast cancer incidence. *Int J Cancer* **119**: 920–924.

Sarkeala T, Anttila A, Forsman H, Luostarinen T, Saarenmaa I, Hakama M (2004) Process indicators from ten centres in the Finnish breast cancer screening programme from 1991 to 2000. *Eur J Cancer* **40**: 2116–2125.

Sarkeala T, Hakama M, Saarenmaa I, Hakulinen T, Forsman H, Anttila A (2006) Episode sensitivity in association with process indicators in the Finnish breast cancer screening program. *Int J Cancer* **118**: 174–179.

StataCorp (2011) Stata Statistical Software: Release 12. StataCorp LP: College Station, TX.

Topo P, Klaukka T, Hemminki E, Uutela A (1991) Use of hormone replacement therapy in 1976-89 by 45-64 year old Finnish women. *J Epidemiol Community Health* **45**: 277–280.

Welch GH, Black W (2010) Overdiagnosis in cancer. *J Natl Cancer Inst* **102**: 605–613.

Wu JC, Hakama M, Anttila A, Yen AM, Malila N, Sarkeala T, Auvinen A, Chiu SY, Chen HH (2010) Estimation of natural history parameters of breast cancer based on non-randomized organized screening data: subsidiary analysis of effects of inter-screening interval, sensitivity, and attendance on reduction of advanced cancer. *Breast Cancer Res Treat* **122**: 553–566.

Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP (2006) Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* **332**: 689–692.

Zahl PH, Jørgensen KJ, Gøtzsche PC (2013) Overestimated lead times in cancer screening has lead substantial underestimation of overdiagnosis. *Br J Cancer* **109**: 2014–2019.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)

APPENDIX 1

Incidence model

Incidence I_{ac} in the 5-year age group a and the 5-year birth cohort c was analysed with the Poisson regression using the following model:

$$ln\, E(I_{ac}) = \mu + \alpha_a + \beta_c + \gamma_{a;\, c=1935-1939}\, I_{a=50-59,\, 60-64;\, c=1935-1939},$$

where

- μ is a baseline, that is, the incidence in the youngest 5-year age group (40–44 years), and the oldest 5-year birth cohort (1920–1924),
- α_a is the effect of other 5-year age groups ($a = 45-49$, 50–54, 55–59, 60–64, 65–69 and 70–74 years) compared with the baseline μ,
- β_c is the effect of other 5-year birth cohorts ($c = 1925-1929$, 1930–1934, and 1935–1939) compared with the baseline μ and
- $\gamma_{a;\, c=1935-1939}$ is the effect of screening in the invited age group 50–59 years ($a = 50-54$ or $a = 55-59$ years) and in the post-screening age group 60–64 years within the cohort (1935–1939), and further
- $I_{a=50-59,\, 60-64;\, c=1935-1939}$ is an indicator divided into two parts

  ○ for $\gamma_{50-59,\, 1935-1939}$, the indicator $I_{a=50-59;\, c=1935-1939} = 1$, if $a = 50-54$ or $a = 55-59$ years; and 0, otherwise
  ○ for $\gamma_{60-64,\, 1935-1939}$, the indicator $I_{a=60-64;\, c=1935-1939} = 1$, if $a = 60-64$ years; and 0, otherwise.

Calculation of the expected incidence rates

Alternative A1

Let us denote that the observed incidence in the last non-invited 5-year birth cohort (1930–1934) is $I^{obs}_{1930-1934}$ and the model-based cohort effect β^{est} between the two youngest 5-year cohorts (1930–1934 and 1935–1939) is $coh^{est} = exp(\beta^{est}_{1935-1939} - \beta^{est}_{1930-1934})$. Then the expected incidence without screening in the invited 5-year birth cohort $I^{exp}_{1935-1939}\, (A1) = I^{obs}_{1930-1934}{}^* coh^{est}$.

Alternative A2

Let us denote that the observed incidence in the invited 5-year birth cohort (1935–1939) is $I^{obs}_{1935-1939}$ and the estimated effect of screening is γ^{est}_{50-59} and γ^{est}_{60-64} in the age groups 50–59 and 60–64 years, respectively. Then the expected incidence without screening in the invited 5-year birth cohort

$$I^{exp}_{1935-1939}\, (A2) = I^{obs}_{1935-1939}{}^* exp(-\gamma^{est}_{50-59}{}^* I_{a=50-59;\, c=1935-1939})^* exp(-\gamma^{est}_{60-64}{}^* I_{a=60-64;\, c=1935-1939}).$$