Effect of organised mammography screening on stage-specific incidence in Norway: population study

Mette L Lousdal*,[1], Ivar S Kristiansen[2], Bjørn Møller[3] and Henrik Støvring[1]

[1]Department of Public Health, Aarhus University, Bartholins Allé 2, Aarhus C DK-8000, Denmark; [2]Department of Health Management and Health Economics, Oslo University, PO Box 1089, Oslo N-0317, Norway and [3]The Cancer Registry of Norway, PO Box 5313, Oslo N-0304, Norway

Background: We aimed to estimate the effect of organised mammography screening on breast cancer stage distribution by comparing changes in women eligible for screening, based on birth cohort, to the concurrent changes in younger, ineligible women.

Methods: In an open cohort study in Norway, which introduced national mammography screening county-by-county from 1995 to 2004, we identified women ($n = 49\,883$) diagnosed with *in situ* or invasive breast cancer (ICD10 codes: D05 or C50) during the period 1987–2011 and born between 1917 and 1980. We estimated relative incidence rate ratios (rIRRs) comparing the development in the screening *vs* historic group to the younger *vs* younger historic group.

Results: Including the compensatory drop, eligible women experienced a 68% higher increase in localised cancers (rIRR = 1.68, 95% confidence interval (CI): 1.51–1.87) than younger women, while the increase in incidence of advanced cancers was similar (rIRR = 1.11, 95% CI: 0.90–1.36). Excluding the prevalence round, eligible women experienced a 60% higher increase in localised cancers (rIRR = 1.60, 95% CI: 1.42–1.79), while the increase in incidence of advanced cancers remained similar (rIRR = 1.08, 95% CI: 0.86–1.35).

Conclusions: Introduction of organised mammography screening was followed by a significant increase in localised and no change in advanced-stage cancers in women eligible for screening relative to younger, ineligible women.

Breast cancer screening programmes have been introduced in most European countries (Giordano *et al*, 2012), yet estimates of their impact on breast cancer mortality vary widely across observational studies, from 0% to 43% (Harris *et al*, 2011; Broeders *et al*, 2012; Irvin and Kaplan, 2014). Screening is intended to advance time of diagnosis to an earlier, more curable cancer stage (Vainio and Bianchini, 2002). A crucial first step in evaluating a screening programme is therefore to identify whether a downward shift in stage distribution is accompanied by a decrease in incidence of advanced stages (Vainio and Bianchini, 2002; Esserman *et al*, 2009; Autier and Boniol, 2012).

Recent studies on stage-specific trends in incidence report a significant increase in early-stage cancers after introduction of screening but only a modest decrease in advanced stages (Esserman *et al*, 2009; Bleyer and Welch, 2012; de Glas *et al*, 2014). Likewise, we found that the gradual implementation of a national screening programme in Norway from 1995 through 2004 was accompanied by a significant increase in the incidence of localised stage cancers. At the same time, the incidence of more advanced stages did not change significantly among women aged 50–69 years when compared with a younger control group ineligible for screening, and trends persisted even when corrected for lead-time bias (Lousdal *et al*, 2014). All previous studies, however, employed an imprecise exposure classification based on trends in mammography screening (Esserman *et al*, 2009), self-reported data (Bleyer and Welch, 2012), or calendar time periods

*Correspondence: ML Lousdal; E-mail: lousdal@ph.au.dk

(de Glas *et al*, 2014; Lousdal *et al*, 2014). A breast cancer mortality study demonstrated the importance of precise exposure classification because a more refined classification revealed a greater reduction in mortality (Olsen *et al*, 2007). Analogously, exposure classification may affect estimates of changes in stage distribution. As the invitation procedure for the Norwegian screening programme is based on birth cohorts, precise exposure classification can be achieved by analysing data according to birth cohorts rather than age groups (Falk, 2014). The use of birth cohorts is also necessary to adequately control for lead-time bias. Lead-time is the interval from the point in time of screening diagnosis to the point in time of clinical diagnosis had the woman not been screened. Effective screening necessarily leads to an initial increase in incidence, because diagnoses are moved forward in time corresponding to the length of lead-time. The increase should later be compensated by a drop in incidence unless the screening-detected cancers represent overdiagnosis. Following birth cohorts ensures that only women with an actual prior invitation to the screening programme will be included, when follow-up is extended to include the expected compensatory drop among women leaving the screening programme (Falk, 2014).

With information on county of residence at diagnosis, we are able to create a more precise classification of actual exposure of birth cohorts to screening invitation during the years in which the national screening programme was gradually introduced in Norway. This allows us to employ a cohort approach, which is methodologically superior to previous, more ecological studies. We aimed to estimate the effect of introducing a screening programme on breast cancer stage distribution by comparing stage-specific incidence in women eligible for screening to the corresponding incidence prior to organised screening relative to the concurrent change in younger, ineligible women.

MATERIALS AND METHODS

Study design and setting. We designed a population-based, open cohort study based on the natural experiment of gradual introduction of organised screening in Norway. Screening started as a pilot project in four counties in 1995–1996 and the remaining counties were subsequently enrolled one by one from 1999 until nationwide coverage in 2004 (Hofvind *et al*, 2007; Kalager *et al*, 2010). Invitation was based on the birth cohort corresponding to women aged approximately 50–69 years during the 2-year screening period (Erztaas, 2003; Hofvind *et al*, 2007). The four start-up counties also invited women born in 1927–28 to the second round in 1998–1999, although they exceeded the upper-age limit (Falk, 2014). The attendance rate was 76% during the period 1996–2005 (Hofvind *et al*, 2007).

Study population. We identified women born between 1917 and 1980 diagnosed with first-time ductal carcinoma *in situ* or invasive breast cancer (ICD10 codes: D05 or C50) during the period 1987–2011 within the Cancer Registry of Norway at the date of diagnosis. We followed cases until death, emigration, or end of follow-up (30 June 2013), whichever came first. The criterion for first-time incidence was that the woman did not appear with an earlier breast cancer in the registry, which includes information on all incident cancers back to 1953 (Larsen *et al*, 2009). We obtained information on stage, tumour size, and residence at date of diagnosis; birth year; and survival after diagnosis. Data on the annual size of the source population were drawn from Statistics Norway. Based on a linear interpolation within each calendar year, we calculated person-time at risk on a monthly basis and assigned it to birth cohorts (see Supplementary Data for details). To comply with ethical requirements for register-based research in Norway, all

data were de-identified by interval censoring the date of diagnosis to the fifteenth day of the month of diagnosis.

Study variables. Invitation to screening, the exposure, was defined by the woman's birth cohort and county of residence at the date of diagnosis (see Supplementary Data for details). Because the date of diagnosis was set to the fifteenth day of a given month, we could not determine whether a diagnosis was made before or after the initiation of screening in the month of implementation. Therefore, we excluded the month of implementation and assigned the following month to be the start of screening. Although screening rounds were defined to last 24 months, in practice the first rounds were prolonged in some counties and shortened in others (Falk, 2014).

For each county, we constructed a unique pattern of four groups: a screening group, a historic group, a younger group, and a younger historic group (Figure 1A). The dotted vertical line indicates implementation date of screening and varies among counties according to the actual month of implementation. The two upper jagged lines indicate eligibility for screening based on birth cohorts, rather than age groups, and 2-year screening rounds. This pattern was replicated in the younger and younger historic groups to ensure balanced groups.

This design defines the group of exposed, that is, the screening group invited to screening and seeks to isolate the effect of introducing a screening programme. The younger and younger historic groups take into account other changes over time that may affect incidence, for example, changes in the level of awareness, breast cancer risk factors, and use of opportunistic screening. The historic and younger historic groups take into account the age effect. When combining all four groups, we estimate the effect that
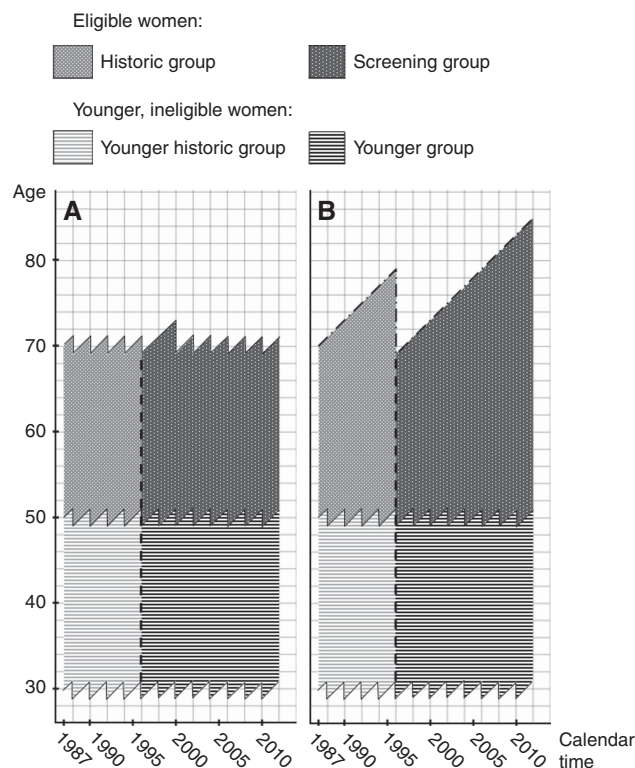

Figure 1. Study design illustrated for the county of Oslo, where screening was introduced on 8 January 1996. Women are eligible for screening based on birth cohorts and programme start. Younger, ineligible women are below the age of screening. (**A**) Follow-up ends at the upper age limit for screening. (**B**) Follow-up of birth cohorts is continued to include the compensatory drop.

could be attributed to the introduction of organised screening. We assume that developments over time in stage-specific incidence in the eligible women (i.e., screening *vs* historic group) and the younger, ineligible women (i.e., younger *vs* younger historic group) would have been similar in the absence of screening.

Outcome is stage at diagnosis as classified by the Norwegian Cancer Registry: *in situ*, localised, regional lymph node involvement, or advanced stages (i.e., local infiltration into skin or chest wall or distant metastatic spread). Secondary outcome is tumour size, classified as *in situ*, diameter $\leqslant 2$ cm, diameter > 2 cm and $\leqslant 5$ cm, diameter > 5 cm, or tumour with infiltration into the skin or chest wall.

Statistical analysis. We tabulated stage-specific incidence, person-years at risk, median age at diagnosis, and median date of diagnosis to describe the study population. Missing data on stage, tumour size, and residence at diagnosis were multiply imputed by chained equations using the *mi* procedures available in Stata 13 (Moons *et al*, 2006; Sterne *et al*, 2009; White and Royston, 2009; White *et al*, 2011). Missing residence was imputed in a multinomial logistic model with stage, tumour size, the Nelson–Aalen estimator of survival, the event indicator (censored at the end of follow-up), age, and date of diagnosis as predictors. Missing stage and tumour size were imputed in a Gaussian normal regression model with stage or tumour size, the Nelson–Aalen estimator of survival, the event indicator, residence, age, and date of diagnosis as predictors and rounded to the closest integer. For comparison, complete case analyses are reported in Supplementary Data.

We conducted the analysis in several stages. We compared stage-specific incidence in the screening group *vs* the historic group and in the younger group *vs* the younger historic group using Poisson regression. We then repeated the analysis with spline-based adjustment for calendar time and age to remove residual confounding within groups. We chose restricted cubic splines to avoid restrictive linearity assumptions and increased the number of knots until associations remained unaffected by adding further knots. We repeated the analysis allowing for interaction between age groups and calendar time, as their potential interaction may be interpreted as cohort effects (Heuer, 1997). As an alternate strategy, we directly substituted the interaction with birth cohort using splines and obtained similar results (not reported, available upon request). For each analysis, the incidence rate ratios (IRRs) for the screening *vs* historic group were compared with the IRRs for the younger *vs* younger historic group to estimate changes specific to the screening group that could be attributed to the introduction of organised screening. We added county as a categorical covariate to the model to explore whether regional differences in stage development changed the IRRs.

To adjust for lead-time, we used the cumulative-incidence method, which continues follow-up of women from the screening and historic groups (Figure 1B) and includes the compensatory drop in the years after a woman has left the screening programme at the upper age limit (Biesheuvel *et al*, 2007). As there is no agreement on the appropriate length of follow-up and some argue in favour of very long follow-up (Falk *et al*, 2013; Njor *et al*, 2013), women were followed for as long as possible, ranging from approximately age 76 to 88 years. In subanalyses, follow-up was truncated at age 74, 79, and 84 years to explore potential dilution of the compensatory drop. As an alternate method to adjust for lead-time, we excluded prevalence rounds at introduction because they are particularly affected by lead-time from prevalent cancers (Biesheuvel *et al*, 2007). In a subanalysis, the prevalence rounds and, in addition, the two youngest cohorts in each round, those invited for the first time, were also excluded.

We conducted various subanalyses to assess the robustness of our results. To explore whether a longer period of follow-up after introduction of screening would affect the incidence of advanced cancers, we repeated the analyses restricted to the four counties

with screening introduction in 1995–1996. To check for uncontrolled confounding, we conducted a negative control analysis in the period 1987–1999 in which we analysed the 12 counties with actual screening implementation after September 1999 as if screening had been introduced in January 1996. A negative finding, that is, relative IRRs around 1, would indicate that our model takes unadjusted confounding into account. We compared the results of this analysis with the concurrent developments in the four counties, which actually did implement screening in 1995–1996. Because of inconsistent registration, we excluded *in situ* cancers diagnosed before 1993 and repeated the analyses. To examine downstaging within stages, we repeated the analyses with tumour size as the outcome (Black and Welch, 1993; Kopans, 2014; Tabár, 2014). Estimates are reported with 95% confidence intervals (CIs) and were deemed significant at a 5% level. All statistical analyses were conducted in Stata, version 13 (StataCorp, College Station, TX, USA).

RESULTS

We identified 50 101 first-time *in situ* or invasive breast cancers among women born between 1917 and 1980 in Norway during the period 1987–2010. We excluded 218 cases diagnosed in the months of implementation. Table 1 provides characteristics of the 49 883 cases included in the analysis. Supplementary Data includes additional stratification based on four groups of counties with introduction of screening at similar points in time. Five cases had missing information on county of residence, 6601 (13%) cases on recorded stage at diagnosis, and 11 669 (23%) on tumour size. Missing information decreased over calendar time with 30, 19, and 7% in the periods 1987–1994, 1995–2002, and 2003–2011, respectively. Missing information varied with 17, 16, 16, and

Table 1. Stage-specific incidence, person-years, age at diagnosis, and date of diagnosis describing the study population excluding five cases with no information on residence in Norway, 1987–2011

	Historic period	Screening period
Younger, ineligible women		
In situ	392	579
I	1229	2223
II	2535	3470
III	205	326
IV	283	232
Unknown	1189	579
Person-years at risk	7 282 618	8 245 954
Median diagnosis age (25th–75th), years	44 (40–47)	44 (40–47)
Median diagnosis date	15 September 1993	15 December 2005
Eligible women		
In situ	459	2668
I	2339	8834
II	3810	6375
III	368	592
IV	600	554
Unknown	2210	1433
Person-years at risk	5 155 249	6 107 811
Median diagnosis age (25th–75th), years	60 (55–65)	59 (54–64)
Median diagnosis date	15 February 1994	15 November 2005
Additional cases with continued follow-up		
In situ	134	142
I	870	839
II	1471	991
III	177	127
IV	288	166
Unknown	984	205
Person-years at risk	1 584 478	1 042 667
Median diagnosis age (25th–75th), years	73 (71–76)	73 (71–76)
Median diagnosis date	15 November 1995	15 October 2008

15% in the four groups of counties with screening introduction in 1995–1996, 1999–2000, 2001–2002, and 2003–2004, respectively.

Overall implementation of screening was associated with a doubling of localised cancers (relative IRR: 1.94, 95% CI: 1.79–2.09), while advanced cancers remained stable (1.00, 0.87–1.15) in eligible women compared with younger, ineligible women (Table 2). Adjusting for age and calendar time, IRRs comparing the screening period with the historic period declined for localised and increased for *in situ* and advanced stages for both eligible and ineligible women, but relative IRRs remained unchanged. Adding an adjustment for the interaction between age and calendar time reduced the size of the relative IRR for *in situ* cancers (2.68, 2.16–3.33) and localised (1.69, 1.52–1.89). For advanced-stage cancers, the relative IRR increased but was statistically insignificant (1.18, 0.95–1.45). Associations remained unaffected when adding county as a covariate (results not shown, available upon request).

With continued follow-up of birth cohorts of previously eligible women, all IRRs comparing the study group with the historic group declined (Table 3). Compared with younger women, eligible women experienced a 68% higher increase in localised cancers (1.68, 1.51–1.87), while the increase in incidence of advanced cancers was similar (1.11, 0.90–1.36). Truncating follow-up at age 74, 79, and 84 years led to nearly identical associations (results not shown, available upon request). The compensatory drop was largest in the model adjusting for potential interaction between age and calendar time. The compensatory drop was more pronounced in stages I and II than in the advanced stages, which showed only a statistically insignificant decrease, when compared with younger women. Figure 2 compares the incidence over time between ineligible and eligible women, including continued follow-up in Oslo. The estimated IRRs correspond to the parallel shifting of the curves at the time of screening introduction.

Excluding the prevalence round, eligible women experienced a 60% higher increase in localised cancers (1.60, 1.42–1.79) than younger women, while the increase in incidence of advanced cancers remained similar (1.08, 0.86–1.35). When, in addition, the two youngest cohorts in each round were excluded, the relative increase in localised cancers persisted (1.47, 1.30–1.66) and increases in advanced cancers remained similar (1.03, 0.81–1.32).

Restricting the analysis to the four counties with earliest introduction of screening and longest follow-up led to similar results. The relative IRR for localised cancers was 1.86 (1.50–2.31) and for advanced-stage cancers 1.25 (0.87–1.80). Using the remaining counties as a negative control showed estimates >1 but with most confidence intervals including 1. When all *in situ* cancers diagnosed before 1993 were excluded, IRRs comparing the study group with the historic group declined, which led to a decrease in the relative IRRs but still indicated a substantial increase in non-invasive cancers compared with younger, ineligible women. Choosing tumour size as the outcome showed a significant 44–57% increase in incidence of tumours ≤2 cm in eligible women relative to younger, ineligible women and a statistically insignificant 23–35% increase in tumours >5 cm (see Supplementary Data for details).

DISCUSSION

Findings. We found a 60–68% increase in localised cancers in women eligible for screening compared with younger, ineligible women when adjusted for cohort, age, and calendar time effects and lead-time. In contrast, the incidence of advanced stages (III and IV) remained stable with a tendency for a higher increase in women eligible for screening relative to younger women.

Table 2. Stage-specific incidence rate ratios comparing the incidence after *vs* before screening introduction and relative incidence rate ratios comparing changes in eligible women with younger, ineligible women in Norway, 1987–2011

Stage	*In situ*	I	II	III and IV	Total
Unadjusted analysis (n^a = 43 489)					
IRR$_{eligible}$	3.40 (3.10–3.73)	2.54 (2.44–2.65)	1.26 (1.21–1.31)	0.80 (0.74–0.87)	1.76 (1.72–1.81)
IRR$_{ineligible}$	1.05 (0.93–1.19)	1.31 (1.23–1.40)	1.10 (1.05–1.16)	0.80 (0.72–0.90)	1.12 (1.08–1.16)
Relative IRR	3.23 (2.77–3.77)	1.94 (1.79–2.09)	1.14 (1.07–1.22)	1.00 (0.87–1.15)	1.57 (1.51–1.64)
Analysis adjusted for age and calendar time (n = 43 489)					
IRR$_{eligible}$	4.02 (3.50–4.61)	2.04 (1.91–2.18)	1.26 (1.18–1.34)	1.18 (1.02–1.37)	1.74 (1.67–1.81)
IRR$_{ineligible}$	1.23 (1.05–1.44)	1.04 (0.96–1.13)	1.08 (1.01–1.16)	1.13 (0.96–1.34)	1.08 (1.03–1.13)
Relative IRR	3.26 (2.79–3.80)	1.97 (1.82–2.13)	1.16 (1.09–1.24)	1.04 (0.91–1.20)	1.60 (1.54–1.67)
Analysis adjusted for age, calendar time, and interaction (n = 43 489)					
IRR$_{eligible}$	3.67 (3.19–4.23)	1.96 (1.83–2.09)	1.27 (1.18–1.35)	1.23 (1.05–1.44)	1.70 (1.63–1.77)
IRR$_{ineligible}$	1.37 (1.13–1.66)	1.15 (1.05–1.27)	1.08 (0.99–1.17)	1.05 (0.87–1.27)	1.13 (1.07–1.19)
Relative IRR	2.68 (2.16–3.33)	1.69 (1.52–1.89)	1.18 (1.08–1.29)	1.18 (0.95–1.45)	1.51 (1.42–1.60)

Abbreviation: IRR = incidence rate ratio.
[a]Number of included breast cancer cases.

Table 3. Lead-time adjusted stage-specific incidence rate ratios comparing the incidence after *vs* before screening introduction and relative incidence rate ratios comparing the changes in eligible women with younger, ineligible women in Norway, 1987–2011

Stage	*In situ*	I	II	III and IV	Total
Continued follow-up (n^a = 49 883)					
IRR$_{eligible}$	3.41 (2.98–3.89)	1.87 (1.75–2.00)	1.16 (1.09–1.24)	1.12 (0.97–1.29)	1.57 (1.51–1.63)
IRR$_{ineligible}$	1.30 (1.07–1.58)	1.11 (1.01–1.23)	1.03 (0.95–1.11)	1.01 (0.84–1.22)	1.06 (1.01–1.12)
Relative IRR	2.62 (2.12–3.24)	1.68 (1.51–1.87)	1.13 (1.03–1.23)	1.11 (0.90–1.36)	1.48 (1.40–1.57)
Excluding prevalence rounds (n = 38 807)					
IRR$_{eligible}$	3.51 (3.01–4.09)	1.79 (1.66–1.93)	1.13 (1.05–1.22)	1.12 (0.93–1.33)	1.55 (1.48–1.62)
IRR$_{ineligible}$	1.47 (1.19–1.83)	1.12 (1.01–1.25)	1.08 (0.99–1.18)	1.04 (0.83–1.29)	1.12 (1.06–1.19)
Relative IRR	2.38 (1.89–3.00)	1.60 (1.42–1.79)	1.05 (0.95–1.16)	1.08 (0.86–1.35)	1.38 (1.29–1.47)

Abbreviation: IRR = incidence rate ratio.
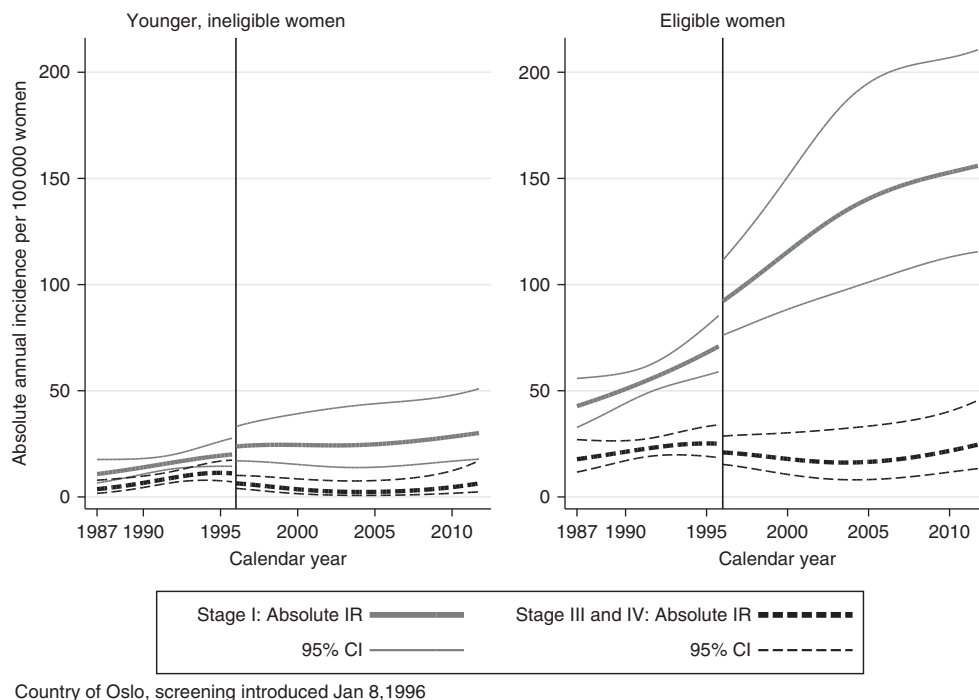[a]Number of included breast cancer cases.

Figure 2. Annual absolute incidence over time in the county of Oslo, 1987–2011. The vertical line indicates introduction of screening.

Strengths and limitations of our study. The main strengths of our study are the follow-up of an entire female population, the high quality data from the Cancer Registry of Norway with 99.95% complete information on breast cancer incidence (Larsen et al, 2009), and the population perspective that considers invitation – and not attendance – as exposure. This effectiveness estimate from a real-world setting is of interest for policymakers and is free of self-selection bias.

The main limitation of our study is the missing information on stage (13% of all tumours) and tumour size (23%) that potentially could introduce selection bias if related to outcome and exposure. However, using multiple imputation to predict missing information yielded findings similar to the complete case analyses, particularly when adjusting for cohort, age, and calendar time effects. In the unadjusted analysis, IRRs comparing the screening period with the historic period dropped when multiply imputed, as missing information was more frequent in the historic period resulting in inflated IRRs. Our imputation model assumes that missing information on stage occurs randomly given other observed patient characteristics, but whether some cases of missing information should instead be considered as a category in itself cannot be determined with available data. Although we do not have individual information on the date of invitation to screening, the use of information on birth cohorts and month of implementation for each county can be expected to be a very close proxy. To assess the reliability of our proxy, we compared it with information on date of invitation among cancer cases, where actual date of invitation is observed and found that 98.6% of invitation dates belonged to the assigned screening group. Information on stage may suffer from time-dependent misclassification owing to changes in classification system. For example, the introduction of TNM classification in 2008 led to an apparent increase in stage III at the expense of stages II and IV. However, we would expect this to apply equally to eligible and ineligible women. Similarly, improved diagnostic techniques may have led to changes over time in tumour classification, known as stage migration (Feinstein et al, 1985). Again, this should be expected to apply equally across eligibility. Of greater concern is downstaging within categories. Our outcome variable may be too crudely categorised to

be sensitive to changes. However, our results with tumour size as an alternative outcome led to identical conclusions.

Temporal changes such as higher awareness and enhancement in imaging technologies may have led to an increase in incidence but would be expected to apply equally to eligible and ineligible women. The extent of opportunistic screening prior to organised screening in Norway has been widely debated (Lynge et al, 2011; Autier, 2013; Weedon-Fekjær et al, 2014). As no substantial increase in incidence of ductal carcinoma in situ was observed until 1996 (Sørum et al, 2010), it seems unlikely that screening has been widespread before the introduction of the programme. Potential opportunistic screening alongside the organised programme in later-starting counties will possibly dilute the initial increase in localised cancers, whereas the drop in incidence of advanced cancers is more likely to be unaffected because it is expected to appear later on corresponding to the length of lead-time. The main confounder remaining is the use of hormone replacement therapy, which reached a peak plateau in 1997–2002 with the greatest consumption among women in the age groups invited for screening (Hofvind et al, 2012). Because screening was gradually introduced, the use of hormone treatment peaked in the screening period in early-starting counties and in the historic period in late-starting counties. There is evidence that tumours among HRT users are smaller (< 2 cm), more often at a localised stage and have a lower proliferation rate (Holli et al, 1998; Manjer et al, 2001). Therefore, use of HRT is an unlikely explanation of the absent decrease in incidence of advanced stages but may partially explain the increase in localised stages in early-starting counties. Using counties without screening as a negative control indicated some uncontrolled confounding, but the estimates had broad confidence intervals, which made it difficult to tell systematic from random variation. This may be due to violation of the underlying assumption of identical but shifted cohort, age, and calendar time effects in the four groups. However, results of the negative control approach were not compatible with a hypothesis of a decreased incidence of advanced breast cancers owing to screening. Because of the open cohort approach, follow-up ends at different ages for each birth cohort and thus cannot include the full compensatory drop for all women. Therefore, the approach with continued

follow-up only partially accounts for lead-time bias. On the other hand, when we excluded the prevalence round, we may have 'over-adjusted' for lead-time as all prevalent cases were excluded. In both analyses, our findings were similar.

Comparison with previous studies. This study resembles the findings from our first paper on trends in stage distribution in Norway (Lousdal *et al*, 2014). There, we did not have information on county of residence at diagnosis and thus had to ignore the implementation period of 1996–2004. In contrast, our present study has a substantially more precise definition of exposure status and thus better approximates the effect at introduction of screening. Moreover, because our previous follow-up model was based on age groups instead of birth cohorts, it included women who had not previously been invited to screening and therefore could not have benefitted. Interestingly, stage-specific IRRs are largely unaffected by this change in methodological approach.

Esserman *et al* (2009) estimated trends for all ages from 1973 to 2006 in the United States and found a substantial increase in localised cancers and a slight decrease in more advanced stages. We used a cohort design *vs* their more ecological approach and are thus in a better position to attribute the changes in stage distribution to the implementation of screening. Similarly, Bleyer and Welch (2012) found a substantial increase in early-stage cancers (including *in situ*) and a small decrease in late stage from 1976 to 2008 in the United Staes in women aged >40 years. To estimate the excess detection from screening, they compared the incidence rates to the incidence among younger women in their so-called best-guess estimate. In contrast, Helvie *et al* (2014) found a substantial reduction in late-stage cancers based on the same US data but with other assumptions about the underlying temporal trends, which have been rebutted (Bleyer *et al*, 2015). We refrain from modelling the underlying background incidence and instead compare with concurrent, younger cohorts from the same counties. The study of de Glas *et al* (2014) is situated in a similar setting with a gradually introduced, nationwide screening programme but focusses on women aged 70–75 years, the existing upper age limit for invitation to screening. They find a strong increase in early-stage tumours without a strong decrease in late stage and therefore question whether the unfavourable effects of screening may outweigh the benefits for the older age group.

Even when we include the compensatory drop or exclude prevalence rounds, the increase in localised cancers persists. Unlike previous studies, we do not find a modest decrease in advanced cancers. This is in line with a review by Autier *et al* (2011) that found no decrease in advanced stages in areas in which mammography screening had been implemented for at least 7 years with a high degree of participation. In contrast to the previous studies, we attempt to adjust for lead-time by employing methods derived from studies on total incidence (Biesheuvel *et al*, 2007). Choosing stage as an outcome *per se* crudely adjusts for lead-time, because the advancement of cases towards earlier stages is rendered visible. However, stage is too crude a categorisation to fully adjust for lead-time. Therefore, we sought to include the compensatory drop, as one would expect fewer advanced cancer in older women previously eligible for screening at the expense of more localised cancers in women eligible for screening. However, when we focussed on the stage-specific compensatory drop, this was more pronounced for stages I and II. This suggests that our understanding of how screening shift stages may need to be reconsidered, possibly owing to substantial heterogeneity in lead-time.

CONCLUSIONS AND POLICY IMPLICATIONS

In this nationwide, population-based study with a refined exposure classification and long follow-up of birth cohorts of women

previously invited to screening, we could not identify a decrease in the incidence of advanced cancers. Our conclusions depend on the plausibility of assuming similar developments over time in eligible and younger, ineligible women. It has been suggested that follow-up of >8–10 years after women have left the screening programme is necessary to fully include the compensatory drop (Falk *et al*, 2013; Njor *et al*, 2013). The underlying assumption is very long lead-times. However, this assumption does not seem to be consistent with the fact that the advanced stages often are the most aggressive, fast-growing tumours. An alternative explanation by Esserman *et al* (2009), which suggests that periodic screening is not capable of detecting the most aggressive cancers early enough, is in agreement with our finding of a negligible compensatory drop in advanced stages.

Future research should investigate stage-specific mortality reductions to achieve a more complete picture of the overall benefits and harms of screening. Furthermore, our population perspective should be supplemented with studies on the stage distribution among participants and non-participants and among interval cancers and screen-detected cancers. The concept of lead-time is intrinsically connected with how screening shifts the stage distribution, and choosing stage as an outcome may thus further our understanding of lead-time heterogeneity.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Autier P (2013) Mammography screening before introduction of the national breast screening programme in Norway. *Int J Cancer* **132**: 1721–1722.

Autier P, Boniol M (2012) Breast cancer screening: evidence of benefit depends on the method used. *BMC Med* **10**: 163.

Autier P, Boniol M, Middleton R, Doré JF, Héry C, Zheng T, Gavin A (2011) Advanced breast cancer incidence following population-based mammographic screening. *Ann Oncol* **22**: 1726–1735.

Biesheuvel C, Barratt 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.

Black WC, Welch HG (1993) Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* **328**: 1237–1243.

Bleyer A, Thomas CR, Baines C, Miller AB (2015) Flawed assumptions used to defend screening mammography. *Cancer* **121**: 320–321.

Bleyer A, Welch HG (2012) Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med* **367**: 1998–2005.

Broeders M, Moss S, Nyström L, Njor S, Jonsson H, Paap E, Massat N, Duffy S, Lynge E, Paci E (2012) The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* **19**: 14–25.

de Glas NA, de Craen AJ, Bastiaannet E, Op 't Land EG, Kiderlen M, van de Water W, Siesling S, Portielje JE, Schuttevaer HM, de Bock GH, van de Velde CJ, Liefers GJ (2014) Effect of implementation of the mass breast cancer screening programme in older women in the Netherlands: population based study. *BMJ* **349**: g5410.

Erztaas AKO (2003) *Kvalitetsmanual: Mammografiprogrammet [Quality Manual: Organized Mammography]*. The Cancer Registry of Norway:

Oslo, Norway. Available from http://www.kreftregisteret.no/Global/Kvalitetsmanualer/kvalitetsmanual_mammografiprogrammet.pdf.

Esserman L, Shieh Y, Thompson I (2009) Rethinking screening for breast cancer and prostate cancer. *JAMA* **302**: 1685–1692.

Falk RS (2014) Why are results of organised mammography screening so difficult to interpret? *Tidsskr Nor Legeforen* **134**: 1124–1126.

Falk RS, Hofvind S, Skaane P, Haldorsen T (2013) Overdiagnosis among women attending a population-based mammography screening program. *Int J Cancer* **133**: 705–712.

Feinstein AR, Sosin DM, Wells CK (1985) The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* **312**: 1604–1608.

Giordano L, Karsa L, von, Tomatis M, Majek O, de Wolf C, Lancucki L, Hofvind S, Nystrom L, Segnan N, Ponti A (2012) Mammographic screening programmes in Europe: organization, coverage and participation. *J Med Screen* **19**: 72–82.

Harris R, Yeatts J, Kinsinger L (2011) Breast cancer screening for women ages 50 to 69 years a systematic review of observational evidence. *Prev Med* **53**: 108–114.

Helvie MA, Chang JT, Hendrick RE, Banerjee M (2014) Reduction in late-stage breast cancer incidence in the mammography era: implications for overdiagnosis of invasive cancer. *Cancer* **120**: 2649–2656.

Heuer C (1997) Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* **53**: 161–177.

Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P (2007) Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. *Eur J Epidemiol* **22**: 447–455.

Hofvind S, Sakshaug S, Ursin G, Graff-Iversen S (2012) Breast cancer incidence trends in Norway–explained by hormone therapy or mammographic screening? *Int J Cancer* **130**: 2930–2938.

Holli K, Isola J, Cuzick J (1998) Low biologic aggressiveness in breast cancer in women using hormone replacement therapy. *J Clin Oncol* **16**: 3115–3120.

Irvin VL, Kaplan RM (2014) Screening mammography & breast cancer mortality: meta-analysis of quasi-experimental studies. *PLoS One* **9**: e98105.

Kalager M, Zelen M, Langmark F, Adami HO (2010) Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* **363**: 1203–1210.

Kopans DB (2014) Study 'guesstimated' how many cancers would have occurred without mammography screening. *BMJ* **349**: g6824.

Larsen IK, Småstuen M, Johannesen TB, Langmark F, Parkin DM, Bray F, Møller B (2009) Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer* **45**: 1218–1231.

Lousdal ML, Kristiansen IS, Møller B, Støvring H (2014) Trends in breast cancer stage distribution before, during and after introduction of a screening programme in Norway. *Eur J Public Health* **24**: 1017–1022.

Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, Lund E (2011) Mammography activity in Norway 1983 to 2008. *Acta Oncol* **50**: 1062–1067.

Manjer J, Malina J, Berglund G, Bondeson L, Garne JP, Janzon L (2001) Increased incidence of small and well-differentiated breast tumours in post-menopausal women following hormone-replacement therapy. *Int J Cancer* **92**: 919–922.

Moons KG, Donders RA, Stijnen T, Harrell FE (2006) Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol* **59**: 1092–1101.

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.

Olsen AH, Njor SH, Lynge E (2007) Estimating the benefits of mammography screening: the impact of study design. *Epidemiology* **18**: 487–492.

Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338**: b2393.

Sørum R, Hofvind S, Skaane P, Haldorsen T (2010) Trends in incidence of ductal carcinoma in situ: the effect of a population-based screening programme. *Breast* **19**: 499–505.

Tabár L (2014) Mammography screening study hampered by inadequate data. *BMJ* **349**: g6823.

Vainio F, Bianchini F (2002) IARC Handbook on Cancer Prevention. Vol. 7. *Breast cancer screening*. IARC Press: Lyon, France.

Weedon-Fekjær H, Romundstad PR, Vatten LJ (2014) Modern mammography screening and breast cancer mortality: population study. *BMJ* **348**: g3701.

White IR, Royston P (2009) Imputing missing covariate values for the Cox model. *Stat Med* **28**: 1982–1998.

White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med* **30**: 377–399.

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