

Why are results of organised mammography screening so difficult to interpret?

Mammography screening has prompted debate in scientific journals and in the media for the past decade. Researchers disagree on the relative benefits and harms of screening. There are a number of methodological challenges associated with evaluating mammography screening. In this article I will describe and discuss the methods used in some published observational studies of overdiagnosis and mortality following the implementation of organised mammography screening in Norway.

The Norwegian Breast Cancer Screening Program (NBCSP) started in four pilot counties in 1995/96 and was gradually expanded in 2004 to include all the counties (1). The primary aim of screening is to reduce breast cancer-specific mortality. Overdiagnosis, defined as breast cancers that would not have been diagnosed in the woman's lifetime unless she had been invited to or had attended screening (2), is considered the greatest disadvantage.

Methodological challenges

There are many methodological challenges associated with evaluating organised screening programmes. They contribute to researchers arriving at different results and conclusions (3–6). The field is characterised by a great deal of disagreement and has been the subject of much debate, also in the Journal of the Norwegian Medical Association (7, 8). I will concentrate here on the challenges associated with the level of detail of the data and the length of the follow-up time in the Norwegian publications. These factors have been commented on previously with respect to some of the studies (9–11), but are systematised here and illustrated with new calculations and visual presentations.

A number of other factors are also deci-

sive for the differing results (3–6). Among others, these include whether the efficacy is evaluated within an intention-to-treat perspective (for invited women) or a per-protocol perspective (for attending women), the assumptions as to what the situation would be without screening, the measure of overdiagnosis applied, whether ductal carcinoma *in situ* (DCIS) is included in the analyses or not, and whether total mortality or breast cancer mortality is studied.

Design and data

Two main types of data are used in the studies: data at group level and data at individual level. In studies using data at group level (ecological studies), there are more limited possibilities to control for confounding factors than in studies where individual data are used as the unit of analysis (12). If summarial figures are used as individual data, it is important to have control of the errors which they may cause.

Women in birth cohorts corresponding to the age range 50–69 years are invited to attend the NBCSP every two years. On the basis of the biennial screening interval and the date when the counties implemented screening (information from the Cancer Registry of Norway), the actual age at the date of invitation varied from 48 to 73.3 years (Falk appendix). To avoid erroneous classification of the women's invitation status (invited/not invited), the data must be analysed on the basis of birth cohort (parallelograms), not age (quadrates), since the women are invited according to birth cohort (Fig 1).

Follow-up

Diagnosis can be advanced by screening (lead-time). An increase in incidence can be expected while the screening is ongoing and a decrease in incidence once screening has ended. Long-term follow-up is necessary, to estimate both overdiagnosis and breast cancer mortality. Studies of overdiagnosis show that a follow-up period of at least ten years after screening ends is necessary to be able to include the compensatory drop in incidence among screened women (4, 5). If the follow-up period is limited, modelling techniques must be resorted to in order to adjust for lead-time.

It takes time from the (first) invitation to screening, to a breast cancer diagnosis, and it takes even more time before women die of the disease. The data from the Cancer Registry show that half the women who were aged 50–69 at the time of diagnosis for clinical breast cancer and who died of the disease in 1991–95, had lived more than 5.5 years with the disease. Since the purpose of screening is to detect tumours at an asymptomatic stage, a longer follow-up period will be required to estimate the effect on mortality of the programme.

Norwegian mortality studies

The results of three breast cancer mortality studies have been published since the implementation of the NBCSP (13–15).

Kalager et al. estimated a decrease of 10% in breast cancer mortality for invited compared with non-invited women (13). They applied the turn of the year closest to the counties' date of implementation as the approximated date of invitation for the women in the invited group. Based on my own calculations, this means that a woman may have been classified as invited up to 2.5 years before receiving her first invitation to the NBCSP (Falk appendix). They analysed the data in relation to age (Fig. 1, quadrate). The women were followed until the end of 2005, which means that the average follow-up time is 2.2 years from the date of diagnosis (13) and, according to my calculations, about 3.5 years from the date of invitation (Falk appendix).

Olsen et al. have calculated the decrease in breast cancer mortality in the four pilot counties as 11% for invited women (14). The five counties that began the programme last were used as control counties, while the ten counties that began in the years 1999–2001 were not included. The women were followed to the end of 2008. This means that the average follow-up time was 5.9 years from the date of invitation.

Hofvind et al. carried out a cohort study in which individual data were based on a precise date of invitation and attendance at screening (15). After correcting for self-selection they calculated that the decrease in breast cancer mortality was 43% for women who attended screening compared with women who did not attend. They esti-

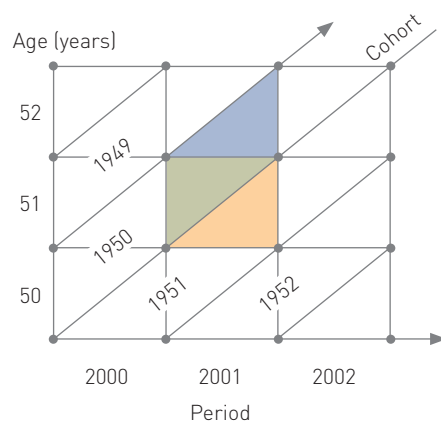


Figure 1 Women are invited according to birth cohort (parallelogram) to attend the Norwegian Breast Cancer Screening Program. Approximation using age (quadrate) will lead to inaccurate results. The example illustrates how women born in 1950 and invited in 2001 cannot be correctly approximated as 51-year-old women

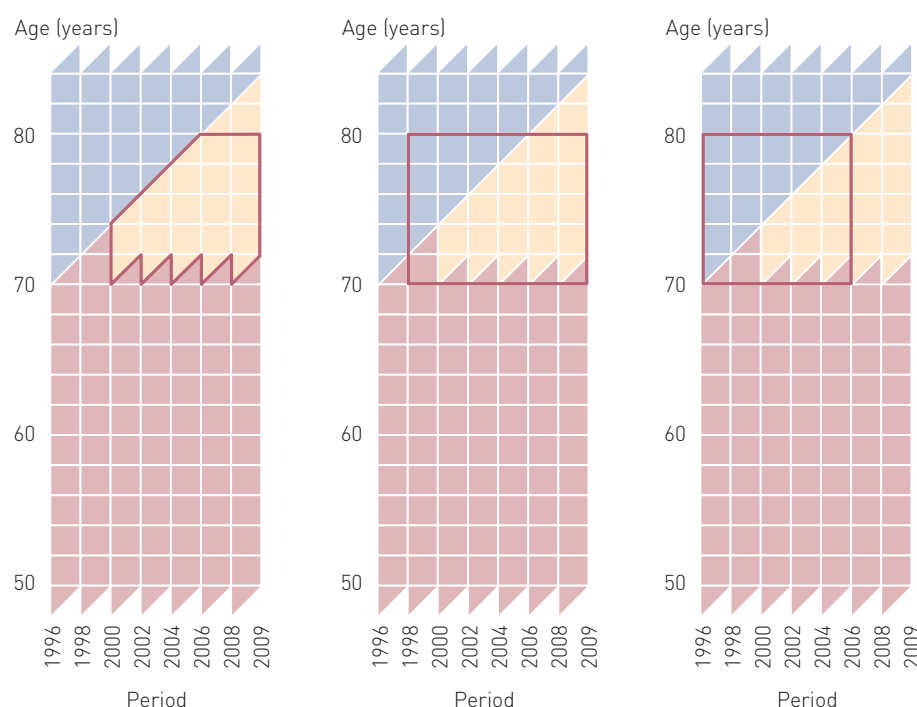


Figure 2 Follow-up after end of screening. Schematic illustration of birth cohorts invited in the four pilot counties of the Norwegian Breast Cancer Screening Program (NBCSP) in the period 1996–2009. The pink-shaded area indicates women who received an invitation. The yellow-shaded area indicates women previously invited. The blue-shaded area indicates women who have never been invited. All three studies evaluating overdiagnosis after the implementation of the NBCSP include women aged 70–79 in the post-screening period (18–20). The red line marks the authors' demarcations. Based on the population data [21] and the start date of the NBCSP in the different counties (information from the Cancer Registry), I have calculated the proportion of women-years within the red demarcation line who have been invited previously (Falk appendix). The study by Falk et al. [20] applies individual data with 100 % previously invited (left-hand panel). Zahl & Mæhlen [18] studied the period 1998–2009 with 52 % previously invited (middle panel). Kalager et al. [19] studied the period 1996–2005 with 29 % previously invited (right-hand panel)

estimated the invitation effect as 36 %. The women were followed to the end of 2010, which gives an average follow-up time of 8.3 years from date of invitation and 5.7 years from date of diagnosis.

Norwegian overdiagnosis studies

Currently, five studies have been published that estimate the extent of overdiagnosis since the implementation of the NBCSP (16–20). The two oldest studies (16, 17) will not be discussed here, as Zahl & Mæhlen have overlapping data and longer follow-up time (18).

Zahl & Mæhlen have performed an ecological cross-sectional study (18). They concluded that 50 % of the breast cancer cases among invited women in the four pilot counties were representative of overdiagnosis. They analysed the numbers in relation to age (Fig. 1, quadrate) up to the end of 2009 (Fig 2, middle panel). My own calculations, based on data from Statistics Norway (21), indicate that 48 % of the women-years have been erroneously included in the post-screening period (Falk appendix). The proportion of overdiagnosis was measured as excess breast cancer cases in relation to cases without screening in the age group 50–69.

Kalager et al. reported that 15–25 % of the breast cancer cases among invited women constituted overdiagnosis (19). They studied the incidence rates of breast cancer among women in the study group compared with incidence rates in three different control groups. They considered the women in relation to age (Fig. 1, quadrate) up to the end of 2005 (Fig. 2, right-hand panel). This means, based on my calculations, that 71 % of the women-years in the four pilot counties have been erroneously included in the post-screening period (Falk appendix). I calculate that this percentage was even higher for the other counties: 81–100 % (Falk appendix). The proportion of overdiagnosis was measured as excess breast cancer cases in relation to cases without screening in the age group 50–79.

We carried out a cohort study using anonymised individual data (20). The women were followed from the date of first invitation and up to the end of 2009 (Fig. 2, left-hand panel). A woman's attendance was classified in relation to her screening history. The proportion of overdiagnosis was estimated for women who follow the national recommendations for ten screenings between the ages of 50 and 69. The time perspective without screening was cal-

culated for age 50 and older. The proportion of overdiagnosis of breast cancer among invited women was estimated as 10–11 %.

There are significant differences between these studies with respect to the accuracy of the data and the length of follow-up. Figure 1 illustrates the inaccuracies which arise if age is used instead of birth cohort in the analysis. In studies of the post-screening period, it is a prerequisite that *all* the women included should have been invited previously, in order to include the compensatory drop in incidence. This condition is not fulfilled for two of the studies (Fig. 2). The measure of overdiagnosis is defined for different women and for different age groups in the three studies mentioned.

Summary

There are many methodological challenges associated with evaluating the NBCSP, and the results must be seen in the light of whichever method is applied. Using summarial figures may lead to erroneous classification of the women's invitation status. It is therefore necessary to apply individual data. Evaluation of mortality and overdiagnosis require long-term follow-up data to provide correct estimates.

This paper is based on a trial lecture for a PhD held at the Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, on 26 September 2013.

See appendix at www.tidsskriftet.no/falkengappendiks

I should like to thank my colleagues and doctoral supervisors Tor Haldorsen, Solveig Hofvind and Per Skaane for their useful comments on this paper.

Ragnhild Sørum Falk
rs@ous-hf.no

Ragnhild Sørum Falk (born 1980) is cand.scient. in biostatistics and took her PhD in 2013 on the thesis *Epidemiological studies of early-stage breast cancer in the Norwegian breast cancer screening program*. She works at the Oslo Centre for Biostatistics and Epidemiology at Oslo University Hospital.

The author has completed the ICMJE form and declares no conflicts of interest.

References

1. Hofvind S, Geller B, Vacek PM et al. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. *Eur J Epidemiol* 2007; 22: 447–55.
2. Day NE. Overdiagnosis and breast cancer screening. *Breast Cancer Res* 2005; 7: 228–9.
3. Marmot MG, Altman DG, Cameron DA et al. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013; 108: 2205–40.

>>>

4. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res* 2013; 15: R41.
5. de Gelder R, Heijnsdijk EA, van Ravestein NT et al. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011; 33: 111–21.
6. Puliti D, Duffy SW, Miccinesi G et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* 2012; 19 [suppl 1]: 42–56.
7. Hofvind S. Organisert mammografiscreening – flere fordeler enn ulemper. *Tidsskr Nor Legeforen* 2013; 133: 619–20.
8. Zahl PH. Informasjonen om mammografiscreening er ikke nøytral. *Tidsskr Nor Legeforen* 2013; 133: 1557–8.
9. Duffy SW, Smith RA. More on screening mammography. *N Engl J Med* 2011; 364: 283.
10. Haldorsen T, Tretli S, Ursin G. Overdiagnosis of invasive breast cancer due to mammography screening. *Ann Intern Med* 2012; 157: 220.
11. Tretli S, Ursin G. Overdiagnostikk ved mammografiscreening. *Tidsskr Nor Legeforen* 2012; 132: 1206.
12. Benestad HB, Laake P. Forskning: Metode og planlegging. I: Benestad HB, Laake P, red. *Forskningsmetode i medisin og biofag*. Oslo: Gyldendal Akademisk, 2004: 83–113.
13. Kalager M, Zelen M, Langmark F et al. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* 2010; 363: 1203–10.
14. Olsen AH, Lynge E, Njor SH et al. Breast cancer mortality in Norway after the introduction of mammography screening. *Int J Cancer* 2013; 132: 208–14.
15. Hofvind S, Ursin G, Tretli S et al. Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. *Cancer* 2013; 119: 3106–12.
16. Zahl PH, Strand BH, Mæhlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004; 328: 921–4.
17. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009; 339: b2587.
18. Zahl PH, Mæhlen J. Overdiagnostikk av brystkreft etter 14 år med mammografiscreening. *Tidsskr Nor Legeforen* 2012; 132: 414–7.
19. Kalager M, Adami HO, Bretthauer M et al. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med* 2012; 156: 491–9.
20. Falk RS, Hofvind S, Skaane P et al. Overdiagnosis among women attending a population-based mammography screening program. *Int J Cancer* 2013; 133: 705–12.
21. Statistisk sentralbyrå. Statistikkbanken. Befolkning. Folkemengden. Tabell 07459: Folkemengde etter kjønn og ettårig alder. www.ssb.no/statistikkbanken/selecttable/hovedtabellHjem.asp?KortNavnWeb=folkemengde&CMSSubjectArea=befolkning&checked=true [10.10.2013].

Received 17 December 2013, first revision submitted 24 January 2014, accepted 6 May 2014. Editor: Siri Lunde Strømme.