

# Norwegian mammography screening – numerous self-contradictions in the evaluation

Public mammography screening was introduced in Norway in 1996. The goal was to reduce mortality from breast cancer by 30 per cent. In 2006, the Research Council of Norway was charged with evaluating the Mammography Programme. The report contains a number of self-contradictions: screening does not cause fewer women to develop neoplasm metastasis, but nevertheless reduces breast cancer mortality.

The 2015 evaluation report concludes that breast cancer mortality has been reduced by 20–30 per cent, and that for every death from breast cancer prevented, five women are overdiagnosed (1).

The incidence of breast cancer has been estimated after adjustment for hormone use

## BOX 1

In this imaginary scenario, which describes 20 years of mammography screening, altogether 550 additional women are diagnosed with breast cancer from 50 to 69 years of age because of mammography screening, compared to a situation in which they had not undergone screening (extra cases). Of these 550, a total of 50 women have received a genuine early diagnosis, whereas 500 have been overdiagnosed (Figure 1).

If when estimating the level of overdiagnosis (the right column in the figure) we assume that 20 per cent of the extra cases are due to an underlying increase in incidence (20 years with an annual increase of one per cent) (16), the number of extra cases is reduced from 550 to 440.

We further assume that women above the screening age (70–79 years) also have a one per cent annual increase in breast cancer incidence. If there were 400 cases in this age group before the introduction of screening, we can expect 480 cases in light of the 20 per cent increase. When we then observe 350 cases in this age group (400 minus the 50 who have received a genuine early diagnosis), these are 130 fewer than those 480 that were expected. If these 130 are subtracted from those 440 extra cases for reasons of early diagnosis, we conclude that 310 women have been overdiagnosed.

By modelling an underlying increase in incidence of one per cent per year while also assuming that there are numerous tumours with long lead times that require follow-up of individuals for ten years after they are no longer invited to screening, the amount of overdiagnosis can be underestimated by approximately 40 per cent.

and an assumed underlying increase in incidence, while the analyses of mortality have not been similarly adjusted. Moreover, it is assumed that there is little or no effect of improved forms of treatment on breast cancer mortality.

Mammography is an x-ray examination of the breasts, and the purpose of mammography screening is to detect tumours while they are still small and localised, thus to enable curative treatment of the women concerned. When a tumour is detected, a biopsy is performed, and approximately one in five turns out to be invasive breast cancer (1). Mammography also assists in detection of a number of tumours restricted to the mammary glands, so-called ductal carcinoma in

situ (DCIS). Some of these tumours develop into invasive breast cancer, but the majority of them do not develop any further (2, 3). Today we are unable to predict which of these tumours have a potential to develop, and which of them will remain unchanged or regress, and the treatment is therefore the same as for invasive breast cancer.

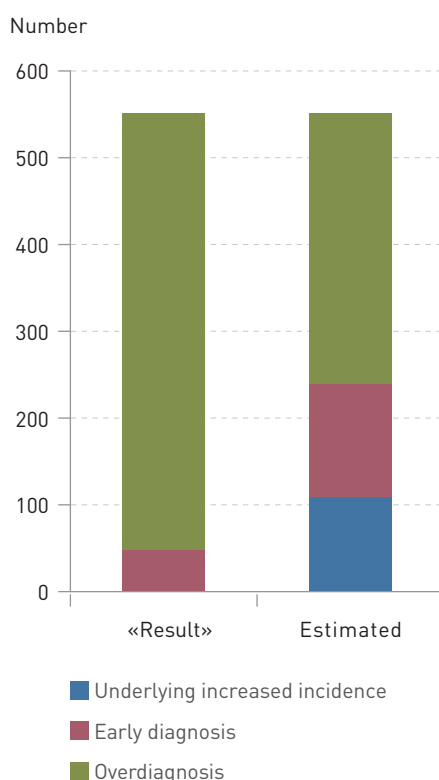
The report from the Research Council of Norway is based on various types of analyses of incidence and mortality from breast cancer. A competitive process resulted in the selection of seven different research groups to conduct the analyses. They were required to use a shared data set consisting of individual data from the Norwegian Cancer Registry and the Causes of Death Registry, linked to a number of other health registries.

## Overdiagnosis

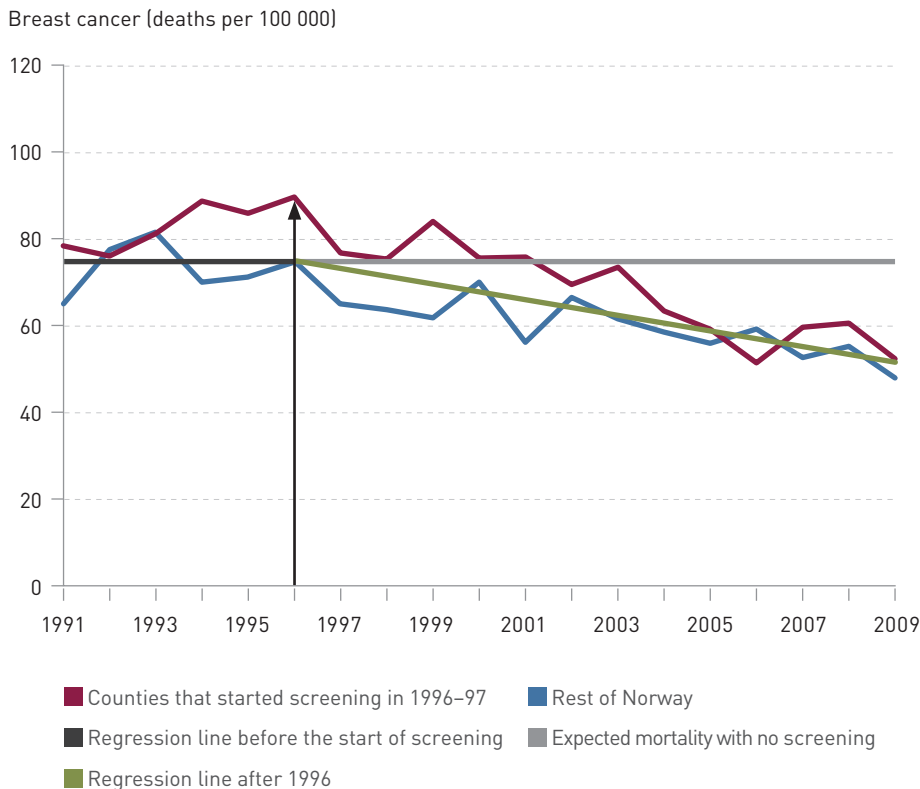
Overdiagnosis means diagnosis of tumours that would otherwise not have resulted in any symptoms during the patient's lifespan (4). Overdiagnosed tumours may include neoplasms that grow very slowly, do not grow and remain subclinical, or disappear spontaneously. Overdiagnosis can easily be estimated in randomised studies, but randomised studies of mammography screenings may no longer be undertaken. Methods have also been developed to estimate the extent of overdiagnosis in public mammography programmes, and a significant difference between the methods is the type of tumours that are included in the denominator. Irrespective of whether we include all cases of breast cancer in the age group 50–69 years (the screening age), 50–74 years or 50–84 years, or whether we estimate the expected number of tumours without screening, the scope of overdiagnosis will vary between 10 and 50 per cent when using the same figure in the numerator (4). While this variability is caused by varying definitions, not bias in the analyses, we will in the following provide three examples of statistical adjustment in the report that in fact increases the risk of bias and underestimation of overdiagnosis.

## Example 1. Adjustment for use of hormones against menopausal symptoms

The report from the Research Council of Norway states that the use of hormones



**Figure 1** Scenario of extra cases of cancer caused by mammography screening. Left column: Of those 550 women who were diagnosed with breast cancer, altogether 50 have received a genuine early diagnosis, while 500 have been overdiagnosed. Right column: If the estimates of overdiagnosis take into account that 20 per cent of the 550 extra cases are caused by an underlying increase in incidence (20 years with a one per cent increase annually) (16), the number of extra cases is reduced from 550 to 440



**Figure 2** The breast cancer mortality in the age group 55–74 years was approximately 76 per 100 000 women before screening was introduced (black arrow), and in the period 2005–09 it amounted to approximately 55 per 100 000. The solid red curve includes the four counties (Akershus, Oslo, Rogaland and Hordaland) that introduced screening in 1996–97, and the blue curve is the rest of Norway. The solid black line is the regression line before the introduction of screening, and the grey line is the expected mortality with no screening. The green line is the regression line after 1996. It shows a reduction of 28 per cent from 1991–95 to 2005–09

against menopausal symptoms was a key reason for the steep increase in breast cancer incidence in Norway during the 1990s, and that it is essential to adjust for the use of hormones with the aid of individual-level data (1).

In 2000, approximately 40 per cent of all post-menopausal women used hormones against menopausal symptoms (1, 5). The report ignores the fact that only half of these women used combination drugs (5), and that only combination drugs increase the risk of breast cancer (5–7). Nor has it been emphasised that the association between hormone use and the risk of breast cancer appears to be 2–4 times stronger in observation studies than in randomised intervention studies (5–7). The likely reason is that hormone use causes the breast to be less penetrable by x-rays, with less contrast between pathological and healthy tissue, and this delays the time of diagnosis (6, 8).

Hormone use is a time-dependent variable, and the start time and duration of the treatment must be included when making the adjustment. Unfortunately, such data were unavailable to the researchers who evaluated the Mammography Programme, and who therefore needed to categorise the women as users, former users or non-users of hormones against menopausal symptoms (5). By cate-

gorising women in this way in observation studies, the risk of breast cancer from hormone use among women is overestimated by several hundred per cent (6, 7). When an excessive part of the increase in breast cancer incidence is explained with reference to hormone use, the estimates of overdiagnosis will be too low.

#### Example 2. Adjustment for an underlying increase in incidence

The report adjusts for an underlying increase in the incidence of breast cancer over time. The incidence of breast cancer increased by approximately one per cent per year for all women in the period 1953–1985. This increase may have been caused by changes in fertility, diet or other lifestyle-related factors. However, the increase may have also been caused by enhanced attention and more opportunistic screening, and thus detection of more small, slow-growing tumours (1, 9–12). After 1985, the incidence increased only in the age group 50–69 years (13), and opportunistic screening most likely plays a significant role in explaining this age-specific increase. If a major proportion of the observed increase in breast cancer incidence can be explained by an underlying increase in incidence, much of the overdiagnosis can be adjusted away in the analyses (Box 1).

**Example 3. Adjustment for long lead times**  
The report claims that short lead times are a key source of error in estimates of overdiagnosis. We take this to mean the number of years of follow-up after women are no longer invited to mammography screening. This will be significant if there are many tumours with a long lead time.

The lead time is the time interval from the detection of a tumour by screening to the point when the tumour would have been detected clinically. For example, women must be followed up until at least age 79 if many of the tumours detected by screening in the age group 65–69 have a lead time of ten years. In theory, a comparison of cumulative rates until age 79 would then be an appropriate method for estimating the extent of overdiagnosis, because there will be only a small possibility of bias. The disadvantage is that the confidence interval around the cumulative rates will grow with the length of the follow-up period (14), causing statistical uncertainty to increase. Long follow-up periods are therefore not necessarily any better than short ones.

It is more serious that the combination of adjustment for long lead times and adjustment for an underlying increase in breast cancer incidence in excess of its real level has introduced a serious bias in the results. In studies of women with tumours that have been diagnosed by mammography but not operated on, the average lead time has amounted to approximately one year (4). Analyses of observational data show the same result (4).

The notion that the average lead time is 2–7 years is based on mathematical models that assume that all tumours grow, and that all increases in incidence found by screening are caused by early diagnosis (i.e. no overdiagnosis is assumed) (4). This self-contradiction – assuming zero overdiagnosis when estimating the amount of overdiagnosis – is completely unreasonable. Moreover, such mathematical models have been falsified in various ways (4, 15). If too long lead times are assumed, the level of overdiagnosis is underestimated (Box 1).

#### Regression of cancer

In a widely quoted article from the Norwegian Mammography Programme, which is not referred to in the report from the Research Council of Norway, we estimated the proportion of cancerous tumours detected by screening that would have disappeared spontaneously (15).

In this study, women who have been invited for mammography screening on three occasions over six years (the test group) are compared to women who have not been screened for four years before being screened once over the subsequent two years. This study has been designed in such a way as to adjust for nearly all differences in risk between the groups, by inclu-

ding the same women in both the test group and the control group.

Without any overdiagnosis, the total number of tumours should be the same in both groups after six years, but we found 22 per cent more tumours in the test group. The study can be interpreted as much of the increase in breast cancer incidence detected by screening includes tumours that would have gone into spontaneous regression had they not been diagnosed by mammography. In addition, nearly all ductal carcinomas in situ must disappear spontaneously or remain unchanged, because surgical treatment does not lead to fewer cases of breast cancer. The purpose of treating and removing carcinomas in situ is to prevent them from developing into cancer at a later stage. Two randomised studies in which women are offered active follow-up versus surgery have been initiated to investigate the regression of ductal carcinoma in situ (2, 3).

### Mortality

Approximately half of all mortality in women aged 50–74 years is due to cancer, but only 6 per cent of this total mortality is caused by breast cancer (6). The report assesses only the effect of mammography screening on breast cancer-specific mortality, not the effect of screening on total mortality or total cancer mortality. The two latter end points are also interesting, since they capture possible increased mortality as a result of cancer treatment (chemotherapy and radiotherapy increase mortality from cardiovascular diseases and other types of cancer) (9).

The report highlights one particular study as being the most credible, and the conclusions on the effect of mammography screening are based exclusively on this (17). While this study concludes with a 28 per cent reduction in mortality from breast cancer, two other studies (18, 19) show a non-significant reduction of approximately 10 per cent – substantial differences in estimates may also be interpreted as reflecting uncertainty as to the real nature of this effect. The preferred study does not adjust for individual risk factors, is based on a statistical method which is described in vague terms and is thus hard to understand as well as not in common use, and the analysis is based on unstated assumptions that have not been validated.

One of the main unstated assumptions is the way in which the effect of better treatment has been modelled. Mortality from breast cancer started to fall in Norway around 1993, i.e. just before the introduction of public mammography screening (Figure 2) (20). Much of this reduction in mortality after 1993 was probably due to the introduction of modern forms of treatment for breast cancer, such as hormone therapy, chemotherapy and trastuzumab (Herceptin). Improved treatment has been

estimated to reduce breast cancer-specific mortality by 30 per cent (21).

The key question is how the effect of improved treatment has been distinguished from the effect of public mammography screening. If it is found that nearly the entire reduction is due to mammography screening, this is an unreasonable finding. A validation analysis has been undertaken, in which pseudo-invitations to screening of women aged 50–69 years in 1990–94 have been added and where a relative risk of approximately one per cent is found (17). Unfortunately, this exercise does not help validate the method.

No mammography screening programmes (10), including the Norwegian one (11), have reduced the incidence of metastatic breast cancer. While the report draws its conclusions exclusively on the basis of mathematical modelling of mortality from breast cancer, a simple observation tells a different story: without any reduction in advanced-stage breast cancer, it is difficult to imagine a reduction in mortality – and nearly all increases in incidence are overdiagnosis.

### Conclusion

The report concludes that the Mammography Programme has caused a reduction in breast cancer-specific mortality of 20–30 per cent, and that for every death from breast cancer prevented, five women are overdiagnosed. This result must be interpreted with caution. The conclusion is based on a narrow selection of available knowledge, and only a few of the end points recommended for evaluation of screening programmes have been included: the authors have not studied whether mammography screening helps detect cancer at an earlier stage or whether total mortality declines. It does not inspire confidence that adjustment for hormone use and an assumed increase in breast cancer incidence are fundamental for the analyses of incidence, but not in the analyses of mortality.

What is most surprising, however, is to see how the effect of improved treatment in the period after introduction of screening has been addressed. Changes in treatment have been estimated to reduce mortality from breast cancer by 30 per cent. It is unreasonable, though, to assume that the treatment effect and the screening effect should both amount to 30 per cent – which would have caused a 60 per cent reduction in breast cancer mortality.

*We wish to thank Henriette Jodal, house officer and PhD candidate, for her valuable input to the manuscript.*

**Per-Henrik Zahl**  
per-henrik.zahl@fhi.no  
**Øyvind Holme**  
**Magnus Løberg**

Per-Henrik Zahl (born 1961), MD, PhD in biostatistics, statistician at the Division of Mental and Physical Health, Norwegian Institute of Public Health. He has published a number of articles about mammography screening and breast cancer mortality.

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

Øyvind Holme (born 1970), PhD and specialist in general internal medicine and digestive diseases at Sørlandet Hospital, Kristiansand, and the Clinical Effectiveness Research Group, Institute of Health and Society, Faculty of Medicine, University of Oslo.

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

Magnus Løberg (born 1979), MD, PhD and post-doctoral fellow at the Clinical Effectiveness Research Group, Institute of Health and Society, Faculty of Medicine, University of Oslo, at the Department of Transplantation Medicine, Oslo University Hospital and at the K.G. Jebsen Centre for Colorectal Cancer Research, Oslo. The author has completed the ICMJE form and declares no conflicts of interest.

### References

1. Research-based evaluation of the Norwegian Breast Cancer Screening Program. Oslo: The Research Council of Norway, 2015.
2. Time Magazine. 2015. <http://time.com/4057310/breast-cancer-overtreatment/> [8.2.2016].
3. New York Times. 2015. [http://nytimes.com/2015/09/29/health/a-breast-cancer-surgeon-who-keeps-challenging-the-status-quo.html?smid=fb-share&\\_r=0](http://nytimes.com/2015/09/29/health/a-breast-cancer-surgeon-who-keeps-challenging-the-status-quo.html?smid=fb-share&_r=0) [8.2.2016].
4. Zahl P-H, Jørgensen KJ, Gøtzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. *Br J Cancer* 2013; 109: 2014–9.
5. Suhrke P, Zahl P-H. Breast cancer incidence and menopausal hormone therapy in Norway from 2004 to 2009: a register-based cohort study. *Cancer Med* 2015; 4: 1303–8.
6. Zahl PH, Mæhlen J. Bias in observational studies of the association between menopausal hormone therapy and breast cancer. *PLoS One* 2015; 10: e0124076.
7. Ioannidis JPD. Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 2005; 294: 218–28.
8. Banks E, Reeves G, Beral V et al. Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. *BMJ* 2004; 329: 477–82.
9. Gøtzsche PC, Jørgensen KJ. Screening for breast cancer with mammography. (review). *Cochrane Database Syst Rev* 2013; 6: CD001877.
10. Autier P, Boniol M, Middleton R et al. Advanced breast cancer incidence following population-based mammographic screening. *Ann Oncol* 2011; 22: 1726–35.
11. Lousdal ML, Kristiansen IS, Møller B et al. Effect of organised mammography screening on stage-specific incidence in Norway: population study. *Br J Cancer* 2016; 114: 590–6.
12. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* 2015; 350: g7773.
13. 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.

>>>

14. Aalen OO. Nonparametric inference for a family of counting processes. *Ann Stat* 1978; 6: 701–26.
15. Zahl PH, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* 2008; 168: 2311–6.
16. Statistisk sentralbyrå. Tabell 08880: Dødsfall, etter kjønn, alder og detaljert dødsårsak (avslutta serie). <https://ssb.no/statistikkbanken/SelectVarVal/Define.asp?MainTable=DodsfallDetaljAld&KortNavnWeb=dodsarsak&PLanguage=0&checked=true> (8.2.2016).
17. Weedon-Fekjær H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study. *BMJ* 2014; 348: g3701.
18. 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.
19. 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.
20. Autier P, Boniol M, Gavin A et al. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. *BMJ* 2011; 343: d4411.
21. Davies C, Godwin J, Gray R et al. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* 2011; 378: 771–84.

*Received 19 February 2016, first revision submitted 16 March 2016, accepted 23 September 2016. Editor: Kaveh Rashidi.*