



# Overdiagnosis in breast cancer screening

Elsebeth Lynge<sup>1</sup>, George Napolitano<sup>2</sup>, Ilse Vejborg<sup>3</sup>, Anna-Belle Beau<sup>2</sup>

<sup>1</sup>Nykøbing Falster Hospital, <sup>2</sup>Department of Public Health, University of Copenhagen, Copenhagen, Denmark; <sup>3</sup>Department of Radiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

*Correspondence to:* Elsebeth Lynge. Nykøbing Falster Hospital, University of Copenhagen, Ejegodvej 63, DK 4800 Nykøbing Falster, Copenhagen, Denmark. Email: elsebeth@sund.ku.dk.

*Comment on:* Pappadis MR, Volk RJ, Krishnan S, *et al.* Perceptions of overdiagnosis of breast cancer among women 70 years of age and older in the USA: a mixed-methods analysis. *BMJ Open* 2018;8:e022138.

Submitted Aug 21, 2018. Accepted for publication Sep 04, 2018.

doi: 10.21037/tcr.2018.09.03

View this article at: <http://dx.doi.org/10.21037/tcr.2018.09.03>

## Introduction

The purpose of breast cancer screening is to reduce breast cancer mortality. In order to achieve this purpose, an X-ray search—a mammography examination—is made of the breast tissue of healthy women to detect potential cancers that have not given rise to symptoms. With early detection, a woman with breast cancer can be offered a more efficient treatment than it would otherwise had been possible. If screening works, women with screen-detected breast cancer will have a better prognosis than women diagnosed with symptoms of breast cancer.

It is an underlying assumption in breast cancer screening that screen-detected cancers would—in the absence of screening—have progressed to symptomatic disease. However, by searching for symptom-free breast cancer in the breast tissue of healthy women, the possibility may arise that breast cancers are detected that would otherwise not have progressed to symptomatic disease in the women's lifetime. These extra cases are called overdiagnosed or overdetected breast cancers. As overdiagnosis is the most commonly used term in the literature, this concept will be used in the rest of this paper. Even in early stages breast cancer is treated with surgery and radiation therapy, and it is a considerable burden on women to undergo breast cancer treatment. Overdiagnosis can therefore have serious consequences. On this basis, overdiagnosis is an important potential, negative side effect of breast cancer screening.

## Premises for understanding overdiagnosis

There are two important premises for the understanding

of overdiagnosis.

### *Epidemiology not biology*

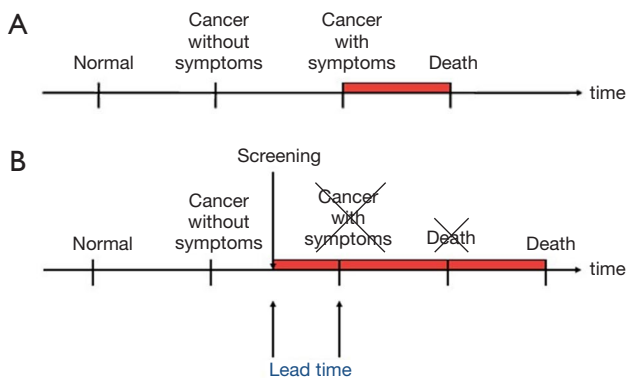
The first point is that an overdiagnosis breast cancer case is a true positive screening result, it is not a false positive result. At the time of screening and assessment of screen positive women, it is not possible to distinguish between potentially progressive and potentially non-progressive—and consequently overdiagnosed—breast cancer cases. If any biology test at the time of diagnosis could distinguish between the two types, this test would of course have been used.

Overdiagnosed breast cancer is therefore not a biological, but entirely an epidemiological phenomenon. It is not possible to say whether or not a given woman had a truly progressive or an overdiagnosed breast cancer. Overdiagnosis can therefore be studied only at the population level.

### *Dynamic of breast cancer incidence rate during screening*

The second point is that breast cancer screening is an intervention on the natural course of the disease, as the diagnoses are made earlier in time than they would have been in the absence of screening. This time period between the time of diagnosis in screening and the time of diagnosis after symptoms is called the lead time (*Figure 1*).

Screening will therefore change the age-specific incidence rate of breast cancer (1,2). This is most easily illustrated from settings where breast cancer screening is organized in biennial rounds say from age 50 to 70.



**Figure 1** Diagnosis of breast cancer without (A) and with (B) screening assuming screening works as intended.

During the first round, there will be cases that would have been diagnosed in the absence of screening and there will be the extra screen-detected cases. This will lead to a sharp increase in the age-specific incidence rate called the prevalence peak (Figure 2). During the subsequent rounds, there will be the screen-detected cases, and due to the earlier detection, these cases will be diagnosed at a younger age than they would have been without screening. This leads to a slight increase in the age-specific incidence rate called the artificial aging. When screening stops, the cases that would have occurred in the absence of screening have been diagnosed already. Therefore new cases are missing.

This leads to a decrease in the age-specific incidence rate called the compensatory dip.

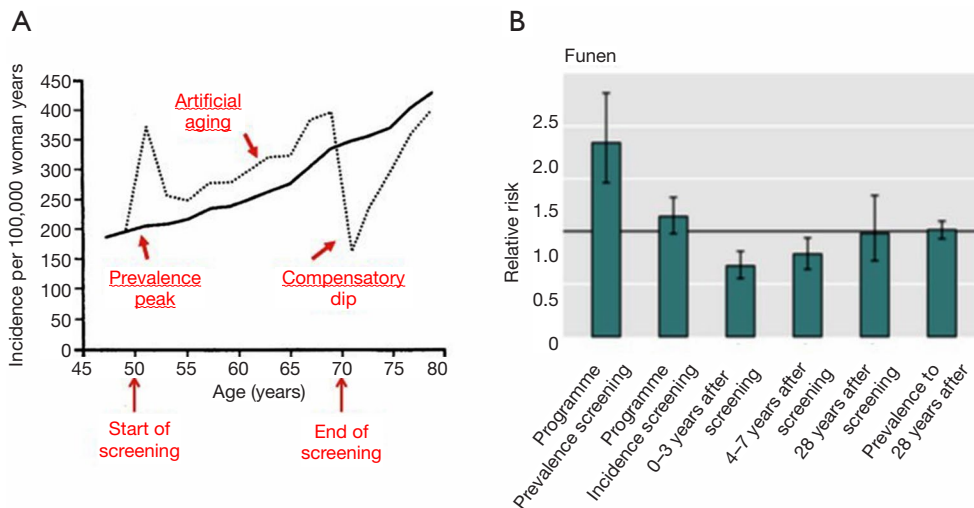
Overdiagnosis occurs if the increases in the incidence rate during the prevalence peak and the artificial aging are not fully compensated by the decrease during the compensatory dip.

**Measuring overdiagnosis**

In order to measure overdiagnosis it is necessary to compare the incidence of breast cancer between a screened cohort of women and an unscreened cohort of women from the age when screening starts until at least 10 years after the end of screening to cover both the prevalence peak, the artificial aging, and the compensatory dip. For women offered screening from the age of 50 to 70 years, this means that for a proper analysis of overdiagnosis, follow-up data from a 30-year period are required from both a screened cohort and from an unscreened comparison cohort of women. Such long-term data are not available, and various proxy methods have therefore been used to measure overdiagnosis.

**Randomized controlled trials (RCTs)**

The RCT is the golden standard for the testing of medical interventions. Breast cancer screening has been tested in a number of RCTs (3), and the decrease in breast cancer



**Figure 2** Breast cancer incidence for unscreened and screened women. (A) Observed incidence of breast cancer in a birth cohort of unscreened women and expected incidence of breast cancer in a birth cohort of screened women (adapted from Boer *et al.*, 1994). Full line: unscreened women; dotted line: screened women; (B) observed relative risk of breast cancer incidence in screened women from Funen County, Denmark, as compared with unscreened women (adapted from Njor *et al.*, 2013).

mortality found in these RCTs forms the basis for the recommendation of screening (4,5).

The RCT has a screened group and an unscreened control group. But in most trials, the unscreened control group was screened at the end of the trial, and the trials did therefore not have long-term incidence data for an unscreened control group. In three RCTs, the control group was not screened at the end of the trial. Based on data from these three RCTs, Marmot *et al.* (6) estimate overdiagnosis to be 11%. However, neither these three RCTs were optimal for the measurement of overdiagnosis. In the Malmö trial from Sweden in the 1980s, women were screened until the average age of 76.5 years leaving insufficient remaining years of life for the compensatory dip to materialize. Two Canadian trials covered women aged 40–49 and 50–59 years, respectively. The control groups were not screened after the end of these trials, but shortly afterwards service screening was offered to all women aged 50–69 years in the majority of the Canadian provinces from which the trial population was recruited (7).

### **Observational data**

In lack of reliable RCT data, several attempts have been made to estimate overdiagnosis from observational data. Denmark offers a particular good possibility for the study of the effect of service screening because population-based screening programs were introduced in two geographical regions up to 17 years before screening was introduced in the rest of Denmark. Denmark at the same time has national population and health data that can be linked via unique personal identification numbers.

We studied overdiagnosis in Denmark using the difference-in-differences methodology on individual cohort data (8). The breast cancer incidence in the screening region during screening was compared with the breast cancer incidence in the same region before screening. To take account of other factors that might have changed the incidence, the comparison from the screening region was adjusted with a similar comparison from the non-screening region. Our analysis showed a prevalence peak, an artificial aging, and a compensatory dip (*Figure 2*). The cumulative breast cancer incidence including both invasive breast cancer and ductal carcinoma in situ in women observed for at least eight years after end of screening was 2.3% higher than expected in the absence of screening. Thus, overdiagnosis was estimated to be 2.3%.

Other researchers have estimated overdiagnosis in

Denmark to be 48.3% of all breast cancer (9), meaning that almost every second screen-detected breast cancer should be overdiagnosed (10). However, this analysis was affected by serious methodological flaws. First, the use of absolute differences in changes over time despite different baseline levels; second, a focus on only non-advanced cancers, and third, an inadequate study design where part of the compensatory dip was calculated based on data from women never invited to screening (10).

The idea of estimating overdiagnosis from analysing advanced and non-advanced cancers separately was first proposed by Welch *et al.* based on data from the United States (US) concluding that in screening, women were more likely to have breast cancer detected “that was overdiagnosed than to have earlier detection of a tumor that was destined to become large” (11). This analysis was based on the assumption that over time “the underlying probability that clinically meaningful breast cancer would develop was stable” given no screening; an assumption not substantiated by pre-screening breast cancer incidence data. Furthermore, the opportunistic and gradually implemented screening in the US does not allow for identification of the prevalence peak, the artificial aging, and the compensatory dip. It is consequently very difficult based on US data to separate out a possible screening effect from the underlying time trend (12).

Some overdiagnosis will inevitably occur in breast cancer screening as some women with screen-detected cancers will die from competing causes of death during the lead time. A modelling study based on data from England & Wales and Norway indicated the inevitable overdiagnosis to be 2–4% (13).

The lack of sufficiently long-term data for the proper measurement of overdiagnosis has left room for a lot a controversy about the size of the phenomenon. Overdiagnosis beyond the inevitable part due to competing causes of death seems to be limited. The most plausible range of overdiagnosis overall was between 1% and 10% in European observational data (14).

### **Perception of overdiagnosis**

#### ***Benefit-to-harm ratio***

As the purpose of breast cancer screening is to decrease breast cancer mortality and as overdiagnosis is the most serious, negative side effect, efforts have been made to weight the two indicators against each other. The benefit-to-harm ratio is the number of prevented breast

cancer deaths divided by the number of overdiagnosed breast cancer cases (15). Using data from the difference-in-differences analysis of individual cohort data from Denmark, we estimated the benefit-to-harm ratio for breast cancer screening in Denmark to be 2.6 for women aged 50 years, invited to screening biennially, and followed until age 79; and to 2.5 for screened women (16). An overall estimate for screened women from European studies was at the same level (15), but the estimate from the Marmot review of RCTs (6) was 0.33 only, and from Norwegian data 0.7 (17). Not surprisingly, the variation in these estimates derived from variation in the estimated overdiagnosis, which is the most difficult component for measure.

The benefit-to-harm ratio is calculated for a given population, and it is therefore first of all a tool for health authorities to decide on offering breast cancer screening or not to this population. In Europe, most health care authorities recommend and provide screening for women in a certain age range, e.g., 50–69 or 50–74 years. According to the US Preventive Services Task Force, screening is recommended for women aged 50–74 years; while the decision to start screening below age 50 should be an individual one; and the current evidence is considered insufficient to assess the benefits and harms of screening in women 75 years and older (5).

### *Personal choice*

As the data on outcome of breast cancer screening derive from the population at large, it has been a challenge for health care authorities to translate these data into information for the individual woman. As a consequence, efforts have been made to develop decision tools to be used in shared decision making between the woman and her care provider primarily for decisions about screening below the age of 50 (18).

In another approach, women aged 74 years were presented with hypothetical cases of outcomes of screening beyond this age, either as a mammogram helping her live to 86 years of age, or as a mammogram finding a slow-growing cancer only (19). Four themes emerged in the women's response to this scenario: (I) resistance to the concept of overdiagnosis; (II) role of the physician's recommendation for screening; (III) confusion with other harms of screening; and (IV) comparison with other health conditions. The authors also found that some women understood overdiagnosis as a concept that applied to populations rather than individuals. We agree with these

women's interpretation, as overdiagnosis is not a biological, but entirely an epidemiological phenomenon. With the presently available diagnostic tools it is not known whether or not a given, newly detected breast cancer will progress or not to become life threatening in the woman's life time. On this basis, we are also sceptical about the authors' idea to "develop interventions that present personal, experiential harms ... differently than harms that reside primarily at the population level" (19).

The effect of screening on breast cancer mortality has not been studied in RCTs in women above the age of 74 years. Nevertheless, the American Cancer Society (ACS) recommends that "women should continue screening mammography as long as their overall health is good and they have a life expectancy of  $\geq 10$  years" (20). The time period of the 10 years after end of screening in the ASC guideline would ensure that women lived through the 10 years of the compensatory dip, and would thus limit the number of diagnosed breast cancers not compensated for. The average life expectancy for US women aged 75 years is at present 13.1 years, and above 10 years for all ethnic groups (21), so according to the ACS guideline they would all qualify for continued screening. As the remaining life expectancy is strongly affected by comorbidity, it has been suggested also to take this factor into account in personalized recommendations on age to stop or continue screening (22). While stratified recommendations can make sense from the point of view of the health care authorities, it might be difficult to implement at the personal level, because most women at the age of 75 years—even those with comorbidity—do not know whether or not they will die within the next 10 years.

### **Conclusions**

It is not possible at the time of screen-detection to know whether a breast cancer will progress or not become life threatening. Overdiagnosis is therefore not a biological characteristic of a cancer that can be tested at the time of diagnosis. It is an epidemiological phenomenon that can be studied at the population level only. Breast cancer screening changes the natural course of the incidence of the disease creating a prevalence peak, and artificial aging, and a compensatory dip. Overdiagnosis occurs if the increase during the first two phases is not compensated by the decrease during the third phase. In order to study overdiagnosis properly, data are needed for a screened cohort and an unscreened comparison cohort for a least

30 years. Such long-term data are not available, and various proxy methods—with very different outcomes—have been used in the study of overdiagnosis. The most reliable data indicate overdiagnosis to account for 1–10% of all incident breast cancer cases, with the inevitably part due to deaths during the lead time constituting a considerably part of this. Overdiagnosis is thus limited given that screened women live at least 10 years after end of screening. This can be used as a guideline at the population level, but it is for most women difficult to translate into personalized recommendations.

### Acknowledgments

*Funding:* None.

### Footnote

*Provenance and Peer Review:* This article was commissioned by the editorial office, *Translational Cancer Research*. The article did not undergo external peer review.

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tcr.2018.09.03>). The authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

### References

- Boer R, Warmerdam P, de Koning H, et al. Extra incidence caused by mammographic screening. *Lancet* 1994;343:979.
- Møller B, Weedon-Fekjaer H, Hakulinen T, et al. The influence of mammographic screening on national trends in breast cancer incidence. *Eur J Cancer Prev* 2005;14:117-28.
- Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening-viewpoint of the IARC Working Group. International Agency for Research on Cancer Handbook Working Group. *N Engl J Med* 2015;372:2353-8.
- Armaroli P, Villain P, Suonio E, et al. European Code against Cancer, 4th Edition: Cancer screening. *Cancer Epidemiol* 2015;39 Suppl 1:S139-52.
- US Preventive Services. Breast cancer screening. Available online: <https://www.uspreventiveservicestaskforce.org/Page/Document/UpdateSummaryFinal/breast-cancer-screening>
- 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.
- Njor SH, Garne JP, Lynge E. Over-diagnosis estimate from The Independent UK Panel on Breast Cancer Screening is based on unsuitable data. *J Med Screen* 2013;20:104-5.
- Njor SH, Olsen AH, Blichert-Toft M, et al. Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ* 2013;346:f1064.
- Jørgensen KJ, Gøtzsche PC, Kalager M, et al. Breast cancer screening in Denmark: a cohort study of tumor size and overdiagnosis. *Ann Intern Med* 2017;166:313.
- Lynge E, Beau AB, Christiansen P, et al. Overdiagnosis in breast cancer screening: The impact of study design and calculations. *Eur J Cancer* 2017;80:26-9.
- Welch HG, Prorok PC, O'Malley AJ, et al. Overdiagnosis in Mammographic Screening because of Competing Risk of Death. *N Engl J Med* 2016;375:1438-47.
- Lynge E, Beau AB, Lophaven S. Impact of assumptions - the example of the Welch-analysis of mammography screening effectiveness. *Acta Oncol* 2017;56:1131-3.
- Falk RS, Hofvind S. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. *Cancer Epidemiol Biomarkers Prev* 2016;25:759-65.
- Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. EUROSCREEN Working Group. *J Med Screen* 2012;19:42-56.
- Paci E. Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate



- of the benefit and harm balance sheet. *J Med Screen* 2012;19:5-13.
16. Beau AB, Lyngé E, Njor SH, et al. Benefit-to-harm ratio of the Danish breast cancer screening programme. *Int J Cancer* 2017;141:512-8.
  17. Hofvind S, Roman M, Sebuødegård S, et al. Balancing the benefits and detriments among women targeted by the Norwegian Breast Cancer Screening Program. *J Med Screen* 2016;23:203-9.
  18. Ozanne EM, Howe R, Omer Z, et al. Development of a personalized decision aid for breast cancer risk reduction and management. *BMC Med Inform Decis Mak* 2014;14:4.
  19. Pappadis MR, Volk RJ, Krishnan S, et al. Perceptions of overdiagnosis of breast cancer among women 70 years of age and older in the USA: a mixed-methods analysis. *BMJ Open* 2018;8:e022138.
  20. Smith RA, Andrews KS, Brooks D, et al. Cancer screening in the United States, 2017: A review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin* 2017;67:100-21.
  21. Arias E, Heron M, Xu J. United States Life Tables, 2014. *NVSS* 2017;66. Available online: [https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66\\_04.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr66/nvsr66_04.pdf)
  22. Cho H, Klabunde CN, Yabroff KR, et al. Comorbidity-adjusted life expectancy: a new tool to inform recommendations for optimal screening strategies. *Ann Intern Med* 2013;159:667-76.

**Cite this article as:** Lyngé E, Napolitano G, Vejborg I, Beau AB. Overdiagnosis in breast cancer screening. *Transl Cancer Res* 2018;7(5):1313-1318. doi: 10.21037/tcr.2018.09.03