Bulliard Jean-Luc (Orcid ID: 0000-0001-9750-2709)
Njor Sisse (Orcid ID: 0000-0003-0429-4176)
Wu Wendy Yi-Ying (Orcid ID: 0000-0002-6169-5155)
Lynge Elsebeth (Orcid ID: 0000-0003-4785-5236)

Breast cancer screening and overdiagnosis

Jean-Luc Bulliard,[1] Anna-Belle Beau,[2] Sisse Njor,[3,4] Wendy Yi-Ying Wu,[5] Pietro Procopio,[6,7] Carolyn Nickson,[6,7] Elsebeth Lynge,[8]

[1] Centre for Primary Care and Public Health (unisanté), University of Lausanne, Lausanne, Switzerland
[2] Pharmacologie Médicale, Faculté de Médecine, Université Paul-Sabatier III, CHU Toulouse, UMR INSERM 1027, Toulouse, France
[3] Randers Regional Hospital, Randers NO, Denmark
[4] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[5] Department of Radiation Sciences, Oncology, Umeå University, Umeå University, Umeå, Sweden
[6] Daffodil Centre, University of Sydney, a joint venture with Cancer Council New South Wales, Sydney, Australia
[7] Melbourne School of Population and Global Health, University of Melbourne, Carlton, Australia
[8] Nykøbing Falster Hospital, University of Copenhagen, Nykøbing Falster, Denmark

Corresponding author:

Dr. Jean-Luc Bulliard
Centre for Primary Care and Public Health (unisanté)
Route de la Corniche 10
Biopôle 2, Bâtiment Proline
1010 Lausanne
Switzerland
Tel: +41 21 314 72 45
Email: jean-luc.bulliard@unisante.ch

Key words: overdiagnosis, estimation, breast cancer, screening, methodology
Word count: 3749
Abbreviations: RCT: Randomized controlled trial

Novelty and impact of the work:

Overdiagnosis is a potential harmful consequence of breast cancer screening. Reliable RCT data on overdiagnosis are not and will not become available so overdiagnosis rates must be estimated from observational or modelling studies. We demonstrate that

observational studies should be carefully designed to avoid methodological pitfalls and provide insights and estimates from well-calibrated modelling studies with different approaches. We conclude that overdiagnosis accounts for at most 10% of breast cancers in targeted women.

ABSTRACT

Overdiagnosis is a harmful consequence of screening which is particularly challenging to estimate. An unbiased setting to measure overdiagnosis in breast cancer screening requires comparative data from a screened and an unscreened cohort for at least 30 years. Such randomized data will not become available, leaving us with observational data over shorter time periods and outcomes of modelling. This collaborative effort of the International Cancer Screening Network quantified the variation in estimated breast cancer overdiagnosis in organized programs with evaluation of both observed and simulated data, and presented examples of how modelling can provide additional insights. Reliable observational data, analysed with study design accounting for methodological pitfalls, and modelling studies with different approaches, indicate that overdiagnosis accounts for less than 10% of invasive breast cancer cases in a screening target population of women aged 50 to 69. Estimates above this level are likely to derive from inaccuracies in study design.

The widely discrepant estimates of overdiagnosis reported from observational data could substantially be reduced by use of a cohort study design with at least 10 years of follow-up after screening stops. In contexts where concomitant opportunistic screening or gradual implementation of screening occurs, and data on valid comparison groups are not readily available, modelling of screening intervention becomes an advantageous option to obtain reliable estimates of breast cancer overdiagnosis.

INTRODUCTION

The purpose of breast cancer screening is to reduce breast cancer mortality. Mammography, as a screening method, allows for detection of breast cancer that has not yet given rise to symptoms. An earlier and more effective treatment can then be offered, and the woman with breast cancer will have a better chance of either being cured, i.e. she will not later die from breast cancer, or live longer before she dies from breast cancer. Both types of outcome will improve survival for women with breast cancer.

It is generally assumed that asymptomatic breast cancers would progress to symptomatic cancers and that screening will find these cancers before symptoms occur. If this is not the case, screening may lead to diagnosis of breast cancer that would in absence of screening not have given rise to symptoms in the woman's lifetime. This is called *overdiagnosis* [1, 2].

It is not possible to know the progression potential of a breast cancer at the time of diagnosis, as we currently have no test that can predict this evolution. All diagnosed women are therefore offered treatment. Breast cancer treatment – even of early stage tumors – involves both surgery and radiation therapy. If a woman is treated for a breast cancer which will not progress to a symptomatic cancer in her lifetime, it is a harm of breast cancer screening, called overtreatment. On this basis, it is important to know the extent to which screened women are under risk of overdiagnosis.

Published overdiagnosis estimates for breast cancer screening have varied substantially [3], causing confusion among women and health professionals, and complicating the overall assessment of the net benefit of breast cancer screening. In this paper, based on the collaborative effort and expertise of the International Cancer Screening Network, we report estimates of breast cancer overdiagnosis based both on observational and modelling studies. We first report estimates of overdiagnosis based on observational data from a unique setting in Denmark to illustrate the degree to which study design and assumptions can influence overdiagnosis estimates. We then present two modelled evaluations of overdiagnosis to demonstrate how, in more typical settings with established screening programs and high quality population level data, modelling can be used to explore specific aspects of overdiagnosis such as the influence of age on the (non)progression of cancer and the association between overdiagnosis and mammographic breast density.

MATERIAL AND METHODS

Screening and breast cancer incidence

The underlying risk of a disease in the population is usually indicated by its incidence rate. However, screening moves the time of diagnosis forward due to earlier detection [4-6]. The lead-time is the interval between the date of diagnosis of a disease (i.e. breast cancer) in screening and the date when this breast cancer would otherwise have been diagnosed after occurrence of symptoms. As the latter date is unknown, the length of the lead-time is unknown, and it is likely to vary across breast cancers. This means the observed breast cancer incidence rate during screening reflects both the underlying risk and the impact of the screening intervention. During the first screen, prevalent asymptomatic breast cancers are detected. We call the "incidence rate" during the first screen the *prevalence peak*, and it can typically be about double the size of the incidence rate given no screening.

During subsequent screens, some incident cases are detected earlier than they would have been without screening. If the underlying risk of disease is constant across age and all cancers diagnosed earlier by screening surface later clinically in absence of screening, the observed incidence rate will return to the level with no screening after lead-time has elapsed. However, the underlying risk of breast cancer increases with age. As a result, a given breast cancer is then diagnosed at an earlier age than would have been without screening. Therefore, the observed incidence rate during subsequent screening will be affected by *artificial aging*, where screening leads to higher incidence in younger age groups.

Finally, when the upper target age range for screening is reached and screening stops, some breast cancer cases that would have occurred in the absence of screening are missing, because they have already been diagnosed during screening. The observed incidence rate will then for some time be lower that it would have been without screening. We call this the *compensatory drop.* Overdiagnosis occurs if the cumulative incidence rate of breast cancer across the prevalence peak, the artificial aging and the compensatory drop exceeds the cumulative incidence rate across this age span in the absence of screening.

Estimating overdiagnosis

While there is general consensus about the numerator in the calculation of overdiagnosis (*i.e*. number of overdiagnosed cases), various denominators have been used [7]. The most frequent being: 1) cumulative incidence during screening; 2) cumulative incidence of screen-detected breast cancers; and 3) cumulative incidence in absence of screening. Therefore, the relative amount of overdiagnosis reported depends strongly on the denominator used in the calculation [7, 8].

Estimating overdiagnosis means quantifying the excess incidence due to breast cancers detected only because of screening. Practically, the frequency of overdiagnosis is virtually impossible to assess, as it would require withholding treatment of diagnosed breast cancers and waiting to determine whether death from other causes preceded symptomatic manifestation of the breast cancers. In principle, it should be possible to calculate the excess cumulative incidence rate of breast cancers from the randomized controlled trials (RCTs) where screening was provided in one arm. However, in most trials, for ethical reasons, the control group was offered screening after the trial ended and the compensatory drop could therefore not be measured by direct comparison of the incidence rates between the intervention and the control groups, but useful modelling studies have been made based on some of these trials [9, 10]. Further, data from the few trials where the control group was not invited to screening after the end of the trial were, for other reasons, not ideal for estimating the compensatory drop [2]. We are thus left with two methods to estimate overdiagnosis: observational studies of excess incidence and modelling studies. Hereafter, we present examples of overdiagnosis for breast cancer from three countries, using observational and modelling methods.

Estimating overdiagnosis from observational data

In order to measure overdiagnosis, we need a screened and an unscreened cohort of women from the age of screening commencement to about ten years beyond the age at which screening stops [11]. For instance, if screening starts at age 50 and ends at age 70, we ideally need data from one unscreened and one screened cohort from age 50 to age 80 (Figure 1 A). Such data are not and will not become available, because it is not realistic to start a screening program for a single birth cohort and wait several decades for the

outcome, while another birth cohort remains entirely unscreened. Therefore, approximations with different lengths of follow-up and denominators have been used to quantify overdiagnosis in observational data, resulting in estimates varying from 1% to 50% [3]. This wide variation is unlikely to reflect true differences in overdiagnosis, as most of the estimates are from European countries, and some of the diverging estimates stem from even the same country. Thus, differences in study design are likely to have had a considerable impact on the size of the estimated overdiagnosis [7, 8].

--- INSERT FIGURE 1 AROUND HERE ---

The Funen county, Denmark, constitutes one of the best settings for an observational study because an organized screening program for women aged 50-69 years started in 1993, while nationwide roll-out of the screening program started only in 2007-2009, and opportunistic screening was rare. Furthermore, in Denmark, the entire population and all health events are registered by personal identification numbers. Individual records on invasive breast cancer were, therefore, available both from the birth cohorts targeted by screening in Funen, from similar cohorts from non-screening counties in Denmark, and from historical control groups from both areas [12] (Figure 1 B). With a follow-up until 14 years beyond the end of screening age, the cumulative incidence in Funen during screening was compared with the cumulative incidence during the same period in non-screening counties, accounting for potential difference in historical incidence trend between the two areas [12, 13].

To assess the impact of study design on estimates of overdiagnosis, the Funen data were also analysed using the designs of five other studies that a) used data from routine screening, b) suggested at least a 20% overdiagnosis, and c) attempted to estimate the compensatory drop by observing women older than the upper age for screening [13]. The Funen data from 1972-2006 were available from Nordcan (https://www.ancr.nu/cancer-data/pc-nordcan/).

Estimating overdiagnosis from modelled data

In numerous countries where concomitant opportunistic screening or gradual implementation of screening occurs, identification of the prevalence peak, the artificial aging and the compensatory drop is often not possible. In some of these contexts, potentially suitable historical or regional control groups exist but data on these comparison groups are not readily available (for instance, Switzerland [14]). In other contexts, historical or regional control groups are simply not possible to find (for instance, the US) (Figure 1 C). In situations with lack of valid comparison groups [15], modelling becomes a key tool for estimating overdiagnosis. Modelling of screening interventions enables follow-up of a fictitious cohort of women throughout their entire lifespan, allowing for competitive death causes, without any screening selection bias. Models also ensure that background incidence rates are perfectly similar in compared groups or over time.

Modelling approaches can be based on a multi-state or a micro-simulation model. Multi-state models use observed data on breast cancer incidence and screening (e.g. screening coverage, sensitivity, etc.) to estimate the natural progression of breast cancer in absence of screening. Micro-simulation models can incorporate detailed tumour natural histories and assume an underlying population breast cancer incidence in the absence of screening, then apply an overlay of screening intervention based on observed cancer incidence and estimated screening behaviour and test accuracy. Breast cancers with long sojourn times are slow-growing or nonprogressive cancers that are more likely to become overdiagnosed cases if detected by screening [9, 10, 16, 17]. Overdiagnosis is estimated by counting simulated invasive screen-detected cancers that would not have become symptomatic before death. Recent extended models also accommodate for nonprogressive lesions and enable precise quantification of the fraction of indolent cancers in settings such as stop-screen trials [18]. Results of recent modelling of overdiagnosis are illustrated thereafter from Sweden [19, 20] and from Australia (http://www.policy1.org/models/breast).

Modelling approach to estimate overdiagnosis of screening in Stockholm county, Sweden

In Sweden, a non-homogeneous multi-state model was developed to address age-specific transition rates and nonprogressive cancers. A hidden Markov model with four latent states and three observed states was constructed to estimate the natural progression of

breast cancer and the test sensitivity (Figure 2). The individual screening history combined with the outcome of screening indicate the subject's observed disease states. The individual observed states are used to construct the likelihood function to estimate the transition rates between disease states and the test sensitivity.

Estimates from this model were previously validated by a simulation study and found to be comparable with the results from the cumulative incidence approach in a randomized trial setting [19]. This method was applied to estimate overdiagnosis in women aged 50-69 years at screening using individual screening data from over 2.3 million women invited by the organized program in Stockholm County over the 25-year period 1989-2014 (overall participation rate: 72.7%; overall recall rate: 2.5%; overall detection rate: 4.9%). The number of overdiagnosed cases was estimated as the expected number of detected nonprogressive invasive breast cancer cases. This number was expressed as proportion of invasive screen-detected breast cancer cases among women invited. Thus, screen-detected cancer cases who died from competing causes before breast cancer became symptomatic are not considered in the estimate.

--- INSERT FIGURE 2 AROUND HERE ---

Modelling approach to estimate overdiagnosis of screening in Australia

In Australia, a continuous-time simulation model of tumour natural history, screening and diagnosis (Policy1-Breast) was developed to evaluate the benefits, costs and harms of different breast cancer screening strategies and, more recently, the impact of COVID-19 on the national screening program and population-level breast cancer [21]. In this stochastic microsimulation model, attributes and behaviours are assigned to individual women, including breast cancer risk based on data prior to the introduction of screening in Australia, tumour progression, age at clinical diagnosis in absence of screening, life-course mammographic density and mortality (breast cancer and other causes). Invasive cancer and DCIS are modelled. A hypothetical screening program is then added to the simulation, replicating observed Australian screening behaviour, including some degree of opportunistic screening outside the program, and modelling screening test sensitivity and specificity according to tumour size and mammographic density.

Policy1-Breast has been developed using 37 years of observed Australian data from 1980 (11 years before the introduction of the breast cancer screening program) to 2016. These data have enabled detailed calibration of the cancer risk, tumour growth and screening components of the model using data from the pre-screening epoch and following the introduction of screening. The design of Policy1-Breast enables modelled estimation of which screen-detected cancers or DCIS are overdiagnosed (*i.e.* would not have been clinically diagnosed in a woman's lifetime), how these outcomes manifest in different sub-groups of simulated women and for different screening protocols, and the size and grade of overdiagnosed tumours at diagnosis. Mammographic density has been modelled using large observed datasets (http://www.lifepool.org/)[22]. Overdiagnosis is expressed in Policy1-Breast as the fraction of all women in the screening age group, regardless of their participation (overall participation rate: ca. 55%).

RESULTS

Observational data

The cohort study indicated no difference in cumulative incidence between Funen and non-screening counties after at least 8 years beyond the end of screening age (mean follow-up of 13 years). These women were 59-70 years at the time of their first screening invitation. Overdiagnosis was estimated to account for 1% (95%CI: -9%;+12%) of invasive breast cancer cases in Funen county. The replication of the design from the five studies with the Funen data led to estimates of overdiagnosis between 13% and 55% [13]. Using the study design applied by Zahl et al. resulted in an overdiagnosis estimate of 41-46% [23]; the ones by Jørgensen and Gøtzsche of 27% and 40-55% [24, 25]; the one by Zahl and Mæhlen of 42% [26]; and finally the one by Kalager et al. of 13-52% [27]. Given that the estimate from the cohort study based on individual records from exactly the same population in the same calendar period was only 1%, the comparison showed that the estimate of overdiagnosis depends strongly on the study design.

Modelled data

Among the 8305 screen-detected cases in the Stockholm county program, the expected number of non-progressive breast cancer cases was 35.9. Overdiagnosis was estimated

to 0.43% (95%CI: 0.10%;2.2%) overall, being 0.87% (95%CI: 0.20%;4.3%) in the prevalent round and 0.31% (95%CI: 0.07%-1.6%) in the subsequent rounds. The non-homogeneous model, which took the age-specific incidence into account, fitted the data better than the simpler, traditional homogeneous model.

In the Australian modelled evaluation, overdiagnosis was estimated to be 11.6% (range 10.9% to 12.3%) of screen-detected cancers and 4.9% (range 4.6% to 5.1%) of population-level invasive breast cancers for women in the historical target age range of the program (50-69 years) over the period 2009-2018 (Table 1). Estimated population-level overdiagnosis was highest among women with low mammographic density at age 50, ranging from 6.2% (5.5% to 7.3%) among women in the lowest mammographic density quintile (Q1) to 3.3% (2.8% to 3.8%) among women in the highest quintile (Q1). Women with higher mammographic density were also estimated to have lower program sensitivity in the screening program, and shorter lead-time at a population level (Table 1).

Key issues and results about estimation of overdiagnosis for breast cancer screening are summarized in Table 2.

DISCUSSION

From an epidemiological perspective, the Funen county provides a highly informative setting for an observational study of overdiagnosis because the incidence in the absence of screening can be reliably estimated from contemporary incidence from non-screening counties combined with control for historical differences, and a sufficiently long follow-up after end of the target screening age is available to account for the compensatory drop. The cohort study design based on Funen data indicated overdiagnosis to account for 1% of breast cancer cases. Using the Funen data, considerably higher estimates of overdiagnosis, of up to 50%, could be reproduced by applying inferior study designs involving unrealistic assumptions about the expected incidence of breast cancer in the absence of screening and underestimation of the compensatory drop due to use of age-group instead of cohort data [13].

These results stress the need for a critical assessment of study design. Two components emerged as particularly important for observational studies. First, how the expected incidence in the absence of screening was estimated and, second, how the compensatory drop was accounted for.

The Funen study design was based on individual records for birth cohorts invited to screening and followed up for up to 14 years after end of screening age. In contrast, all the study designs selected for replication used an age-period design whereby the studied birth cohorts changed from one calendar year to the next. In these studies, the incidence of breast cancer in the absence of screening was mainly estimated either with linear extrapolation of the prescreening incidence for women of screening age or from the incidence trend in women below screening age [13]. Both methods are prone to biases as, for instance, the incidence in the absence of screening has in fact increased more than a linear regression would indicate. Furthermore, it is difficult in the age-period design to account correctly for the compensatory drop, and most of the selected studies ended up including data from some women above screening age who were never invited for screening. This will inevitably lead to an underestimation of the compensatory drop, thus an overestimation of overdiagnosis.

The difference in modelled estimates of overdiagnosis among screen-detected cancers in Stockholm county and in Australia (0.4% vs 4.9%) may reflect variations between the two models. It has been demonstrated in other contexts that variations in outcome can occur when different modelling techniques or assumptions are applied to the same data set. Estimates of the relative contribution of screening to the reduction in breast cancer mortality in the United States between 1975 and 2002 varied from 28% to 65% across seven models [28]. Estimates of the number needed to screen to prevent one breast cancer death in the UK population aged 55-79 years ranged from 64 to 257 across four scenarios in major reviews of mammographic screening [29]. For overdiagnosis specifically, when flawed estimates based on unrealistic assumptions are corrected, the estimated amount of overdiagnosis for Swiss screening programs dropped from 25% to 3% [30]. However, these changes may also be due to differences in health settings. The Australian estimate of 11.6% population-level overdiagnosis for women aged 50-69 is similar to previous

modelled estimates of 9.2% (95%CI: -0.2%;14.7%) and 10.2% (95%CI: -5.2%;19.7%) for Australian women aged 50-59 and 60-69, respectively [31].

The finding from the Australian model that overdiagnosis increases with decreasing breast density is logical since screening test sensitivity is modelled to be higher for women with lower breast density, so that small, slow growing cancers, which are more likely to be overdiagnosed, are more likely to be detected in less dense breasts. The longer lead-time for less dense breasts would be mitigated to some degree by the reporting period commencing in a steady-state screening program so that higher lead-time in less dense breasts would also shift some of their overdiagnoses to before 2009. The association between breast density and overdiagnosis warrants further investigation, particularly in terms of how targeted supplemental or alternative imaging tests for women with higher mammographic density such as those currently being trialled might inadvertently increase overdiagnosis (ClinicalTrials.gov Identifiers: NCT03672331 (MyPEBS), NCT01315015 (DENSE), NCT04097366 (BRAID)).

A common problem for observational studies of overdiagnosis in service screening programs is determining breast cancer incidence in the absence of screening, as there is no ideal unscreened control group. Predictions such as age-period-cohort modelling of pre-screening incidence rates can be sensitive to biases, as occurrence of new risk factors can affect the breast cancer incidence in a way that cannot be captured by modelling of old rates; this was seen for instance after the rapid uptake of hormone therapy in Norway [32]. The risk profile of breast cancers may also change over time, as seen in the Netherlands, where the proportion of women with dense breast increased across birth cohorts [33]. Modelling studies, while no replacement for observed data, are potentially unaffected by this problem as each simulated individual also acts as its own control, all confounders being perfectly equal.

If prediction of breast cancer cases alone is affected by a similar inaccuracy, this could have important consequences for the study of overdiagnosis. In cohort studies where non-screened cohorts were available as control groups and the follow-up was sufficiently long to account for the compensatory drop, overdiagnosis has been estimated to be below

10%, e.g. Funen 1% [15]; Copenhagen, Denmark 3% [15]; and Helsinki, Finland 5-7% [34]. The 10-15% difference between predicted and observed numbers for cancer cases overall could indicate that the true rate of overdiagnosis is too small to be captured correctly in studies that have to rely on pre-screening data for prediction of incidence rates and/or model calibration [35].

CONCLUSION

Overdiagnosis and the associated overtreatment are important potential harmful consequences of breast cancer screening. It is, however, a very difficult phenomenon to study because it ideally requires data from a screened and a non-screened cohort for at least 30 years. As such data are not and will not become available, we are left with observational data from shorter time periods and with outcomes of modelling. The present collaborative effort of the International Cancer Screening Network quantified the variation in estimated overdiagnosis with evaluation of both observed and modelled data, and presented examples of how modelling can provide additional, valuable insights. Modelled estimates of increasing overdiagnosis with decreasing breast density warrants further study in the context of ongoing trials targeting supplemental or alternative imaging tests for women with dense breasts. Although our study showed some variation in the estimates of overdiagnosis, the most reliable observational data, corroborated by two examples of modelled data, indicate that overdiagnosis accounts for at most 10% of invasive breast cancer cases in the target population aged 50-69 years for screening. Estimates above this level are likely to derive from inaccuracies in study design.

Data availability statement

Observational data used for this study are publicly available from the Nordcan website https://www.ancr.nu/cancer-data/pc-nordcan. Dataset from Sweden and the script of estimation of overdiagnosis is made available upon reasonable request and permissions granted by data/script contributors. The Policy1-Breast model is currently protected intellectual property. Further information is available from the corresponding author upon request.

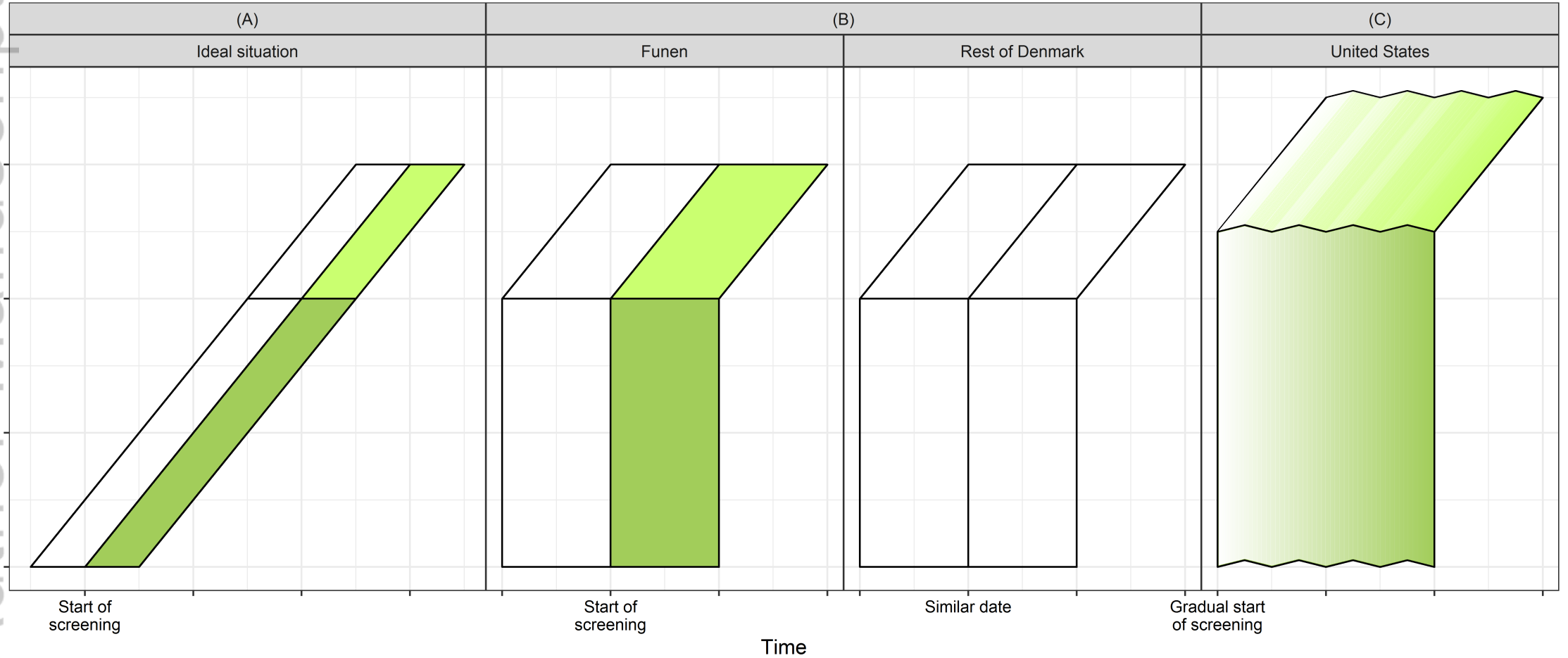
Conflict of interest

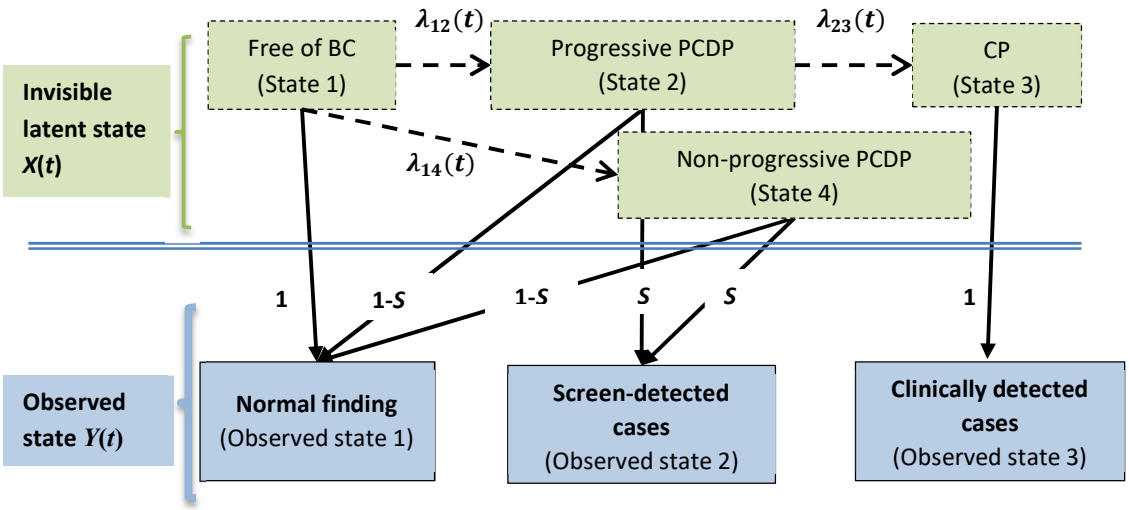
The authors do not report any conflict of interest.

REFERENCES

1. Bulliard JL, Chiolero A. Screening and overdiagnosis: public health implications. *Public Health Rev* 2015;**36**: 8.
2. Lynge E, Napolitano G, Vejborg I, Beau AB. Overdiagnosis in breast cancer screening. *Transl Cancer Res* 2018;**2018**: 1-6.
3. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, Paci E, Euroscreen Working Group. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* 2012;**19 Suppl 1**: 42-56.
4. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol* 2007;**8**: 1129-38.
5. Boer R, Warmerdam P, de Koning H, van Oortmarssen G. Extra incidence caused by mammographic screening. *Lancet* 1994;**343**: 979.
6. Moller B, Weedon-Fekjaer H, Hakulinen T, Tryggvadottir L, Storm HH, Talback M, Haldorsen T. The influence of mammographic screening on national trends in breast cancer incidence. *Eur J Cancer Prev* 2005;**14**: 117-28.
7. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011;**33**: 111-21.
8. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer* 2013;**108**: 2205-40.
9. Gunsoy NB, Garcia-Closas M, Moss SM. Modelling the overdiagnosis of breast cancer due to mammography screening in women aged 40 to 49 in the United Kingdom. *Breast Cancer Res* 2012;**14**: R152.
10. Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, Myles JP, Warwick J. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* 2005;**7**: 258-65.
11. 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.
12. Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E. Overdiagnosis in screening mammography in Denmark: population based cohort study. *BMJ* 2013;**346**: f1064.
13. Njor SH, Paci E, Rebolj M. As you like it: How the same data can support manifold views of overdiagnosis in breast cancer screening. *Int J Cancer* 2018;**143**: 1287-94.
14. Bulliard J-L, Ducros C, Jemelin C, Arzel B, Fioretta G, Levi F. Effectiveness of organised versus opportunistic mammography screening. *Ann Oncol* 2009;**20**: 1199-202.
15. Ripping TM, Ten Haaf K, Verbeek ALM, van Ravesteyn NT, Broeders MJM. Quantifying Overdiagnosis in Cancer Screening: A Systematic Review to Evaluate the Methodology. *J Natl Cancer Inst* 2017;**109**.
16. Yen MF, Tabar L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer* 2003;**39**: 1746-54.
17. Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J* 2006;**12**: 338-42.
18. Ryser MD, Gulati R, Eisenberg MC, Shen Y, Hwang ES, Etzioni RB. Identification of the fraction of indolent tumors and associated overdiagnosis in breast cancer screening trials. *Am J Epidemiol* 2019;**188**: 197-205.

19. Wu WY, Nystrom L, Jonsson H. Estimation of overdiagnosis in breast cancer screening using a non-homogeneous multi-state model: A simulation study. *J Med Screen* 2018;**25**: 183-90.
20. Wu WY, Tornberg S, Elfstrom KM, Liu X, Nystrom L, Jonsson H. Overdiagnosis in the population-based organized breast cancer screening program estimated by a non-homogeneous multi-state model: a cohort study using individual data with long-term follow-up. *Breast Cancer Res* 2018;**20**: 153.
21. Nickson C, Procopio P, Deij S, Velentzis L. COVID-19 scenario modelling for cancer screening programs. Cancer Council NSW, Cancer Research Division, 2020.
22. Kavanagh AM, Byrnes GB, Nickson C, Cawson JN, Giles GG, Hopper JL, Gertig DM, English DR. Using mammographic density to improve breast cancer screening outcomes. *Cancer Epidemiol Biomarkers Prev* 2008;**17**: 2818-24.
23. Zahl PH, Strand BH, Maehlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* 2004;**328**: 921-4.
24. Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* 2009;**339**: b2587.
25. Jorgensen KJ, Zahl PH, Gotzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC women's health* 2009;**9**: 36.
26. Zahl PH, Maehlen J. Overdiagnosis of breast cancer after 14 years of mammography screening. *Tidsskrift for den Norske laegeforening : tidsskrift for praktisk medicin, ny raekke* 2012;**132**: 414-7.
27. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. *Ann Intern Med* 2012;**156**: 491-9.
28. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer EJ. Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* 2005;**353**: 1784-92.
29. Duffy S, Chen T, Smith R, Ming-Fang Yen A, Tabar L. Real and artificial controversies in breast cancer screening. *Breast Cancer Manage* 2013;**2**: 519-28.
30. de Koning HJ, Heijnsdijk EA. Swiss Medical Board Mammography screening predictions for Switzerland: importance of time-periods. *J Med Screen* 2015;**22**: 201-6.
31. Beckmann K, Duffy SW, Lynch J, Hiller J, Farshid G, Roder D. Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects. *J Med Screen* 2015;**22**: 127-35.
32. Weedon-Fekjaer H, Bakken K, Vatten LJ, Tretli S. Understanding recent trends in incidence of invasive breast cancer in Norway: age-period-cohort analysis based on registry data on mammography screening and hormone treatment use. *BMJ* 2012;**344**: e299.
33. Napolitano G, Lynge E, Lillholm M, Vejborg I, van Gils CH, Nielsen M, Karssemeijer N. Change in mammographic density across birth cohorts of Dutch breast cancer screening participants. *Int J Cancer* 2019;**145**: 2954-62.
34. Heinavaara S, Sarkeala T, Anttila A. Overdiagnosis due to breast cancer screening: updated estimates of the Helsinki service study in Finland. *Br J Cancer* 2014;**111**: 1463-8.
35. Moller B, Fekjaer H, Hakulinen T, Sigvaldason H, Storm HH, Talback M, Haldorsen T. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Stat Med* 2003;**22**: 2751-66.





Novelty & Impact Statement: IJC-20-3285.R1

The detection of breast cancer before symptoms arise greatly increases the chance of prolonging survival or even curing malignancy. However, the assumption that asymptomatic disease progresses to symptomatic disease is a major factor in breast cancer overdiagnosis. While estimates of overdiagnosis vary substantially, the present analysis of observational data and data from modeling studies shows that overdiagnosis accounts for less than 10 percent of invasive breast cancer cases among women ages 50 to 69. The findings reaffirm the idea that observational studies require careful design to avoid methodological pitfalls and highlight the value of insight gained from well-calibrated modeling studies.

Table 1: Estimates of overdiagnosis, program sensitivity and lead-time by quintile of mammographic density at age 50 in breast cancer screening for women aged 50-69 years in Australia, 2009-2018 (mean (min-max) over ten simulations).

Population quintile of breast density at age 50	Overdiagnosed cancers (% of invasive screen-detected cancers)		Overdiagnosed cancers (% of invasive cancers, population level)		Program sensitivity (in %)		Lead time of screen-detected cancers (years)	
Q1	12.9	(11.2-14.9)	6.2	(5.5-7.3)	84.6	(83.3-87.7)	4.1	(3.7-4.5)
Q2	12.3	(11.0-13.2)	5.7	(5.1-6.2)	82.5	(80.8-83.9)	3.9	(3.6-4.3)
Q3	12.1	(10.2-15.4)	5.4	(4.5-6.8)	79.4	(77.5-82.9)	3.9	(3.6-4.3)
Q4	11.3	(8.5-13.7)	4.5	(3.5-5.5)	73.5	(72.0-75.2)	3.5	(3.0-4.2)
Q5	9.7	(8.0-10.9)	3.3	(2.8-3.8)	65.6	(63.1-68.1)	3.1	(2.8-3.5)
All women	11.6	(10.9-12.3)	4.9	(4.6-5.1)	76.1	(75.0-77.1)	3.7	(3.5-3.9)

Table 2: Key points and results about estimation of overdiagnosis in breast cancer

- Overdiagnosis is detection at screening of breast cancer that would otherwise not have become clinically manifest in the woman's life time
- Overdiagnosis is the potentially most negative side effect of breast cancer screening
- Biologically, it is not possible at time of diagnosis to distinguish between potentially lethal and non-lethal breast cancers
- Overdiagnosed cases cannot (yet) be identified; overdiagnosis can only be measured at population level
- Observation of overdiagnosis requires an unrealistic long follow-up time to be captured in randomized controlled trials
- Overdiagnosis can be estimated from natural experiments or from modelling
- Many attempts to estimate overdiagnosis from observational data have suffered from methodological errors
- Estimation of overdiagnosis in modelling implies some untestable assumptions
- Use of appropriate methods in both observational and modelling studies indicate that the estimated amount of overdiagnosis in breast cancer screening of women aged 50-69 years varied much less than previously reported and constitutes at most 10% of incident breast cancers