Overdiagnosis in the Dutch and Norwegian breast cancer screening program

Paula van Luijt

Overdiagnosis in the Dutch and Norwegian breast cancer screening program

Paula van Luijt

ISBN: 978-94-6299-463-8

Overdiagnosis in the Dutch and Norwegian breast cancer screening program

Thesis, Erasmus University

©Paula A. van Luijt

All right reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the author or copyright-owing journals for previously published chapters.

Cover:Martirio di Sant'Agata van Francesco Guarino (1611-1651)

The studies reported in this thesis were primarily funded by the National Institute for Public Health and the Environment, RIVM. For any grant pertaining to specific studies the reader is referred to the individual papers published in their respective journals.

Overdiagnosis in the Dutch and Norwegian breast cancer screening program

Overdiagnose bij borstkanker screening

Stand van zaken en een internationaal perspectief

Proefschrift

ter verkrijging van de graad van doctor

aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 29 november 2016 om 11:30 uur

door

Paula Afke van Luijt

geboren te Den Haag

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotor:	Prof.dr. H.J. de Koning
Overige leden:	Prof.dr. C. Verhoef
	Prof.dr. J.W.W. Coebergh
	Prof.dr. J.A. Roukema
Copromotor:	Dr. E.A.M. Heijnsdijk

CONTENTS

CHAPTER 1:

GENERAL INTRODUCTION

BREAST CANCER

Breast cancer is as old as history itself, the first description of breast cancer was found on the Edwin Smith papyrus from 1600 BCE.(1) Before the 20th century it was considered an unspeakable disease. Mostly as a result of the hard work of female laymen advocates people and doctors became more aware of the disease and treatment options were developed.(1) Nowadays breast cancer patients are united in patient associations, there is a breast cancer awareness month, and major advances in treatments have been made.

Breast cancer is a form of cancer that arises from breast tissue. Most breast cancers originate from glandular tissue (adenocarcinoma). All cancers start off as a single cell, as the cell divides and multiplies the new cells form a tumor. Initially the tumor is confined to a natural compartment; this is called an in situ carcinoma. In breast cancer these are mostly ductal carcinoma in situ (DCIS), or less commonly lobular carcinoma in situ (LCIS). As the tumor grows and develops, it may break through the basement membrane, then it is called invasive disease (Figure 1).(2) It is not certain if in situ carcinoma is an obligatory precursor of invasive cancer. With many diagnoses of breast cancer, only invasive disease is found, without evidence of in situ carcinoma. The size of the tumor and its relationship to surrounding tissues, possible lymph node involvement and distant metastases determine its stage.

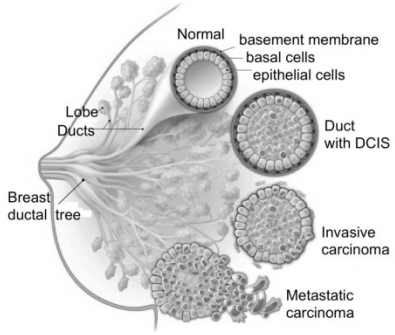


Figure 1. Different stages of breast cancer, in situ carcinoma and invasive carcinoma.

Breast cancer is most commonly classified using the TNM classification (Table 1). T stands for tumor, N for lymph nodes, and M for metastases. Invasive cancer (T1 to T4) can spread via lymph nodes or via the bloodstream (hematogenic). Lymph node metastases manifest first in the first draining lymph node (sentinel node). If there are lymph node metastases, the disease is node positive. Hematogenic spread occurs in lungs, pleura, brain, skin, bones, liver, adrenals or peritoneum (metastatic disease). Disease with distant metastases is considered incurable.

Table 1. TNM classification in breast cancer.(3)

Stage		
DCIS		Ductal carcinoma in situ
T1		Tumor 2 cm or less in greatest dimension
	T1a	0,1 cm or more and less than 0,5 cm in greatest dimension
	T1b	0,5 cm or more and less than 1 cm in greatest dimension
	T1c	1 cm or more and less than 2 cm in greatest dimension
T2		2 cm or more and less than 5 cm in greatest dimension
T3		5 cm or more in greatest dimension
T4		Tumor of any size with direct extension to chest wall and/or to skin (ulceration or skin nodules)
N0		No regional lymph node metastasis
N1		Metastasis in movable ipsilateral level I, II axillary lymph node(s)
N2		Metastasis in ipsilateral level I,II axillary lymph node(s) that are clinically fixed, or in ipsilateral internal mammary lymph node(s) in the absence of axillary lymph node metastasis
N3		Metastasis in ipsilateral infraclavicular (level III axillary) lymph node(s) with or without level I, II axillary lymph node involvement or internal mammary lymph node involvement.
M0		No distant metastasis
M1		Distant metastasis

TREATMENT

Treatment of DCIS usually consists of a local excision followed by radiation therapy. Complete mastectomy is considered for large tumors, or with multi-centricity of disease. Staging of axillary lymph nodes is only indicated if women are younger than 55, if the mammogram shows a mass, if the histopathology is suspect of invasive disease, or if histopathology shows an intermediate or poorly differentiated DCIS.(4)

In the treatment of small invasive disease local excision is possible if a cosmetically acceptable result is expected and if proper loco-regional tumor control can be obtained. In all other cases treatment consists of mastectomy. Local excision is usually followed by radiation therapy. For small invasive disease, the axillary lymph nodes are evaluated. Dissemination does not need to be examined, unless per-operative findings or postoperative histopathology shows a higher disease stage.(4) Adjuvant systemic therapy is considered for all patients with positive lymph nodes and patients with unfavorable characteristics (young age, negative receptors, or large tumors). Neo-adjuvant systemic therapy (given before surgery) is considered for locally advanced disease.(4)

SURVIVAL

In recent decades in most developed countries the mortality rate of breast cancer has been declining, despite increasing incidence.(5) In the Netherlands 5-year survival has increased from 77% in cancers diagnosed between 1989 and 1993 to 87% in cancers diagnosed between 2008-2012.(5)

The increased survival is probably a combined result of advancements in treatment and early detection. Trials studying the use of adjuvant chemotherapy were conducted since 1995. A review in 2005 showed a reduction in breast cancer death by about 35% for first generation adjuvant chemotherapy regimens. Since then breast cancer mortality reduction has increased further with 20% for second generation regimens and again with 20% with third generation regimens.(6)

Some of the improved survival may be attributable to the fact that more breast cancer will be detected at an earlier stage, the period of time between diagnosis and death becomes inevitably longer. Relative survival is expected to be better. However, also the absolute number of breast cancer deaths has reduced from 3,365 in 1989 to 3,014 in 2014.(5)

SCREENING TRIALS AND UK INDEPENDENT PANEL REVIEW

In 1948 Jacob Gershon-Cohen demonstrated the feasibility of detecting occult carcinomas with X-ray. Fourteen years later a large study was conducted by Egan et al. in which two-view mammograms were read and the readers were able to distinguish the benign conditions from the malignant conditions without clinical examination.(1)

Between 1963 and 1991 the following randomized trials were conducted to evaluate mammography screening: New York Health Insurance Plan (HIP), Malmö I and II, Swedish two county trial (Kopparberg and Östergötland), Canada I and II, Stockholm, Göteborg, UK age trial, and Edinburgh. These trials compared women invited to screening to women not invited to screening. Most of the trials found that among the women invited for screening, fewer women died of breast cancer. Treatments as well as screening methods have improved since the trials. Therefore the effect of screening can be different nowadays.

Because there has been a lot of debate on the benefits of screening, an independent research panel has been established in the UK in 2013 performing an extensive review of the evidence for benefits and harms of breast cancer screening. They found that for 10,000 women invited to screening from age 50 for 20 years 43 breast cancer deaths will be averted.(7)

More recently the International Agency for Research on Cancer assessed the cancer-preventative and adverse effects of different screening methods. They found there is sufficient evidence that mammography screening reduces breast cancer mortality for women aged 50-74 years, and this reduction outweighs the risk of radiation-induced breast cancer. They also found sufficient evidence that mammography screening induces overdiagnosis.(8)

EFFECT EVALUATION OTHER THAN TRIALS

To evaluate the effect of screening, several case-control studies have been conducted. Case-control studies compare the number of women who died of breast cancer to the number of women who did not die from breast cancer with regards to exposure to screening. Otto et al found that participating in breast cancer screening reduces the risk of dying of breast cancer with 49%.(9)

Cohort studies examine the risk of breast cancer death in a cohort of women exposed to screening, compared to a cohort not exposed to screening. Cohorts can be historical, if the cohorts are defined by a difference in the time at which both cohorts were exposed to screening.(10) Cohorts can also be geographical, if a certain county or district has been exposed to screening, and the other (possibly neighboring) cohort has not been exposed to screening. (11) Kalager compared historical groups and found an overdiagnosis rate of 15-25%.(10) Autier compared neighboring countries, that had implemented mass breast cancer

screening at a different time and suggested that screening did not play a direct part in the reductions in breast cancer mortality.(11) The difficulty with historical and geographical cohorts is the introduction of bias. Historical cohorts suffer from bias because treatments, screening strategies and overall survival change over time. Geographical cohorts suffer from bias because neighboring counties or countries can have differences in the accessibility of health care, or the use of opportunistic screening.

SCREENING IN THE NETHERLANDS

After a successful initial pilot mammography screening program in Utrecht and Nijmegen starting in 1974, screening was gradually implemented between 1989 and 1997.(12) Every woman aged 50-69 years old receives an invitation for screening every other year. The invitation is accompanied by an information flyer, which contains information on the harms and benefits of screening and emphasizes the voluntary nature of the program. In 1998 the program was extended to include women aged 70-74.

All examinations are read by two radiologists. If there is a difference in assessment, the two readers can confer with one and other to reach consensus, or call in a third party for arbitration. The radiologists that reports the latest assessment, decides.

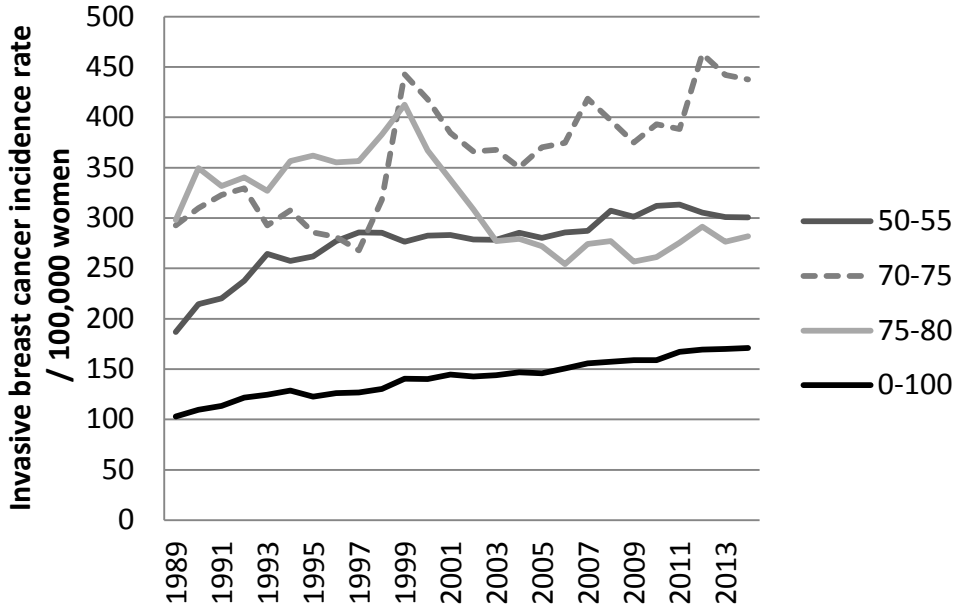
The program made a transition from screening with film screen mammography to digital mammography between 2004 and 2009. Since 2009 all screening examinations are digital. Digital mammography screening has a higher recall rate, resulting in a higher detection rate, particularly for DCIS and small invasive carcinoma. The positive predictive value of a recall is lower for digital mammography. One major advantage of digital mammography screening is that it is easier to compare the images to previous screens, and that the work-flow on the computer is much easier than it was with old-fashioned mammograms, which had to be hanged on the lightbox before evaluation.

Initially all women had a two-view examination (cranial caudal and mediolateral oblique) at first screening, and only a one-view mammogram (mediolateral oblique) at every subsequent screening, unless the technician saw a need to perform a second view. Reasons for performing a second view mammography included: an abnormality not seen on the previous examination, technical difficulties that might impair interpretation, or patients with specific symptoms. A study on two-view mammography in 2012 found that routine two-view mammography at subsequent screenings may modestly increase cancer detection at an earlier stage, with limited additional screening costs.(13) After 2010 all women routinely had a two-view mammography at all screening examinations.

BREAST CANCER INCIDENCE IN THE NETHERLANDS

Breast cancer incidence in women aged 50-55 has risen steeply in the Netherlands since the introduction of population-based screening in 1990. This steep increase is the result of the first screening rounds when many prevalent breast cancers were detected. After the initial steep increase, the increase became much less steep (Figure 2a).(5) In women aged 70-75 we see that breast cancer incidence peaks in 1999, immediately after the program was expanded in 1998. The breast cancer incidence in women aged 75-80 decreases after the expansion of the program, because most tumors detected at this age have now been diagnosed in the age group 70-75. DCIS incidence increased with the introduction of screening in women aged 50-55 in 1990, and with the expansion of the program in 1998 in women aged 70-75. DCIS incidence gradually increases in women of all ages and women aged 75-80 (Figure 2b).

A



B

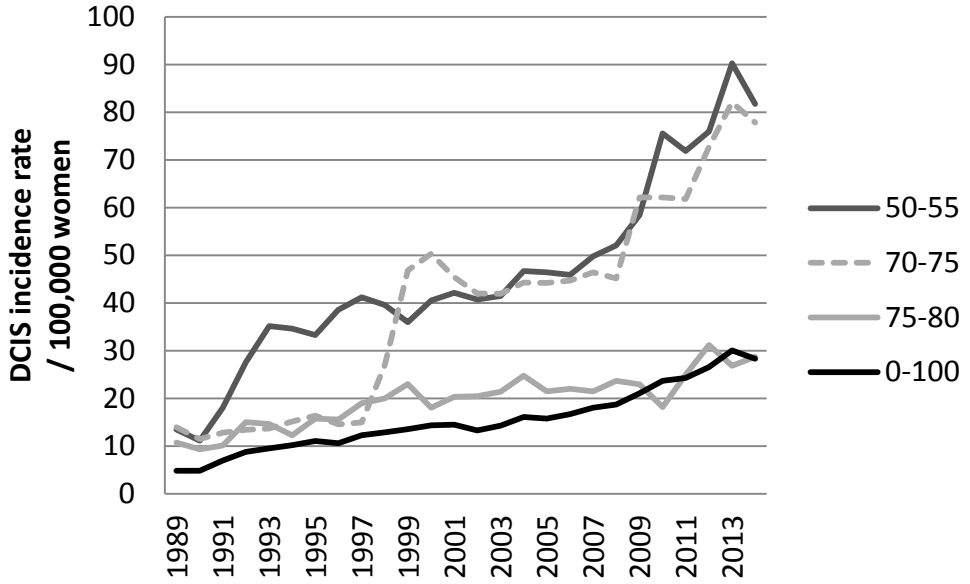


Figure 2. A: Breast cancer incidence per 100,000 women in the Netherlands from 1989 (the last year before population-based screening was introduced) to 2014 (the most recent year with complete data) by age group. B: DCIS incidence per 100,000 women in the Netherlands from 1989 (the last year before population-based screening was introduced) to 2014 (the most recent year with complete data) by age group.

BREAST CANCER MORTALITY IN THE NETHERLANDS

Breast cancer mortality in women over 50 has been declining since the early 1990's (Figure 3).(5) This is the result of both population-based screening and the introduction of adjuvant therapy; chemo therapy and hormonal therapy.(14)

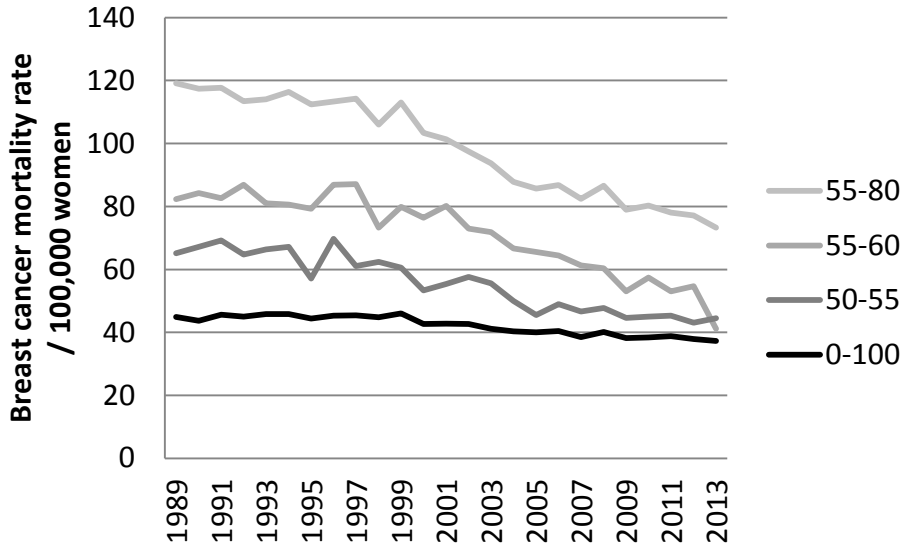


Figure 3. Breast cancer mortality per 100,000 women in the Netherlands from 1989 (the last year before population-based screening was introduced) to 2013 (the most recent year with complete data) by age group.

NORWAY

The Norwegian Breast Cancer Screening Program (NBCSP) has been gradually introduced between 1996 and 2005. The NBCSP has been the topic of a fierce debate on the benefits and harms of a population-based screening program for breast cancer.(10, 15) For this reason, and because the program has been implemented a decade later than the Dutch program, we applied to participate in the recent evaluation of the NBCSP that was conducted by the Research council of Norway.(16)

BREAST CANCER INCIDENCE IN NORWAY

In Norway breast cancer incidence in women aged 50-54 has risen spectacularly since 1995 (Figure 4).(17) This is remarkable, since population screening was gradually introduced in Norway between 1996 and 2005.(18) Breast cancer incidence in women aged 70-74 declines in 1999 because these women are no longer screened, and the tumors that would have been detected at this age, have been detected earlier by screening.

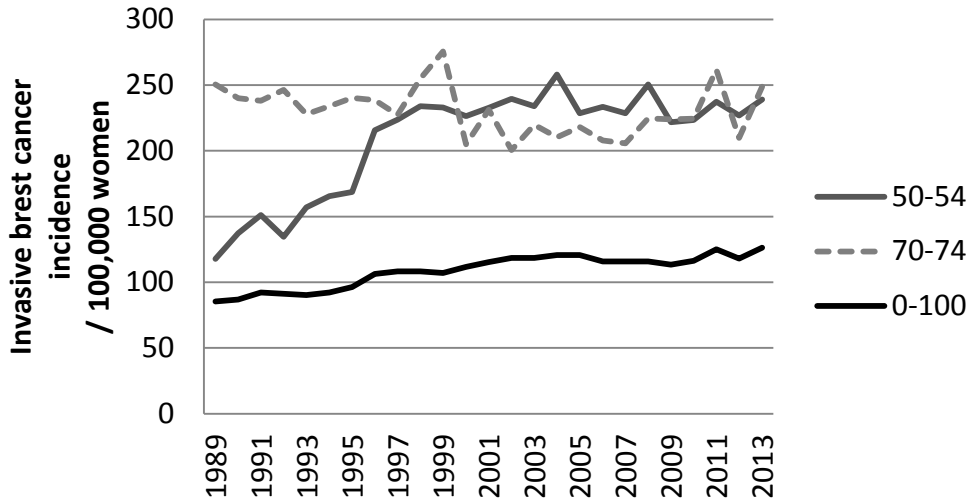


Figure 4. Invasive breast cancer incidence in Norway per 100,000 women from 1989 to 2013, by age group.

BREAST CANCER MORTALITY IN NORWAY

Breast cancer mortality in Norway has been declining in women aged 55-80 from the early 1990's. (Figure 5).(17) This decline occurred prior to the introduction of population-based screening. This is probably the result of adjuvant therapy, greater awareness, and the use of (opportunistic) screening.

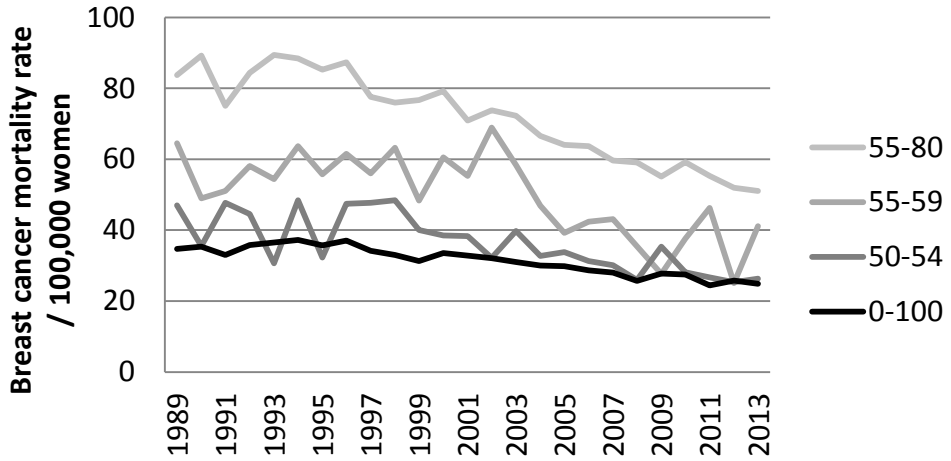


Figure 5. Breast cancer mortality in Norway per 100,000 women from 1989 to 2013, by age group.

SCREENING IN NORWAY

In Norway, population-based screening was gradually introduced between 1996 and 2005.(18) Compared to other Nordic European countries this is relatively late. Sweden started in 1986, Finland in 1987, and Denmark in 1991.(19) All women aged 50-70 years old are invited to mammography screening every other year. All women routinely have a two-view mammogram. All examinations are read by two radiologists. The mammogram is taken in a mobile unit, or in a dedicated stationary unit.(16) Because Norway is a vast country with low population density, distances to a screening unit are larger than in the Netherlands, and travel costs contribute considerably to the cost of screening.(20)

DISADVANTAGES OF SCREENING

The aim of population-based screening for cancer is to reduce the cancer specific mortality. Screening prevents cancer specific mortality by detecting cancer early; at a stage the disease can still be cured. The disadvantage of early detection is that screening also detects disease that would never have become apparent if a woman was never exposed to screening. This is called overdiagnosis. (21) Overdiagnosis occurs when a woman dies of other causes than breast cancer prior to the time when the disease would have become manifest. This can be the result of an untimely death, or because the disease in itself has an indolent character.

DCIS, or small invasive cancer may grow very slowly, or DCIS may even never become invasive disease. Overdiagnosis is the major cause of harm in population-based screening.

Because it is impossible to know on an individual level which diagnosed disease is overdiagnosis, all women with a diagnosis of breast cancer will be treated accordingly. Thus overdiagnosis leads to overtreatment. The treatment of an overdiagnosed disease cannot contribute to life years gained, since the patient will die of other causes, regardless of the breast cancer diagnosis.

CALCULATION OF OVERDIAGNOSIS

Breast cancer incidence in women of the screening age is higher than in women not exposed to screening. This is the result of early diagnosis. After women reach the upper age limit for screening, breast cancer incidence should drop, since cancers that would have occurred in women of this age, were already detected by screening earlier in life and were already treated. The higher incidence in women of the screening age is called excess incidence. The lower incidence in older women is called deficit.

The easiest way to estimate overdiagnosis is to subtract the deficit incidence (after the upper age limit of screening has been reached) from the excess incidence (in women of the screening age). For a population this can be done by subtracting the incidence without screening in women of all ages from the incidence with screening in women of all ages (Figure 6).

Figure 6. Theoretical breast cancer incidence per 100,000 women in women of all ages without screening and with screening by age. Graphical depiction of the excess incidence and the deficit, these are the areas between the curves.

With regards to the definition of overdiagnosis the UK Independent Review Panel proposed two definitions:

1. from a population perspective: the number of excess diagnoses minus the number of deficit diagnoses divided by the total number of diagnoses in invited women of all ages.

2. from an individual perspective: the number of excess diagnoses minus the number of deficit diagnoses divided by the total number of diagnoses in invited women of the screening age.

They estimate that overdiagnosis rate for a population invited to screening is about 11%, and for an individual invited to screening is about 19%. These estimates are based on data from the trials performed in the 1980's.

We used the definitions of overdiagnosis put forward by the UK Independent Review Panel throughout the thesis to allow comparability.

METHODS TO ESTIMATE OVERDIAGNOSIS

To calculate overdiagnosis we need an estimate of the cancer incidence in the absence of screening. Because population-based screening is now implemented in most countries, these data can no longer be observed. Women choosing not to participate in population-based screening are not a reliable control group since their decision not to participate can be correlated with their breast cancer risk.(22)

There are several methods to estimate incidence rate in the absence of screening. Some authors have used breast cancer incidence prior to the introduction of screening.(10, 15) Comparing this incidence to the current breast cancer incidence may introduce bias. The underlying background incidence of breast cancer may very well be different today than it was twenty or even ten years ago.

Another method is to use a model. An age period cohort model studies a population and mathematically derives the individual effect of age of the patients, period of study, and birth-cohort.(23) The aim is to eradicate bias as a result of age, period or cohort. A microsimulation model such as the MIcro-simulation SCreening ANalysis (MISCAN) models the individual life histories of all women in a population.(24)

MISCAN

The Micro-simulation SCreening ANalysis (MISCAN) model generates life histories of all women in a fictional population. The model predicts the onset of cancer, the detection of cancer, and the time and cause of death. Cause of death can be breast cancer, or other causes. A screening scenario can be superimposed on these life histories and may alter the time of detection of cancer, and the cause and time of death. All of the parameters can be calculated in a situation with screening, and in a situation without screening. The model can be calibrated to mimic different populations and different screening strategies.

Population composition is based on demographic data on population composition (number of women of a certain age). The microsimulation model calculates the complete life history for each individual woman in the model. Each woman is given a date of birth, a date of death, and a cause of death, based on national statistical data. The probability of having an onset of breast cancer is determined by the incidence of breast cancer in the population of interest, prior to the introduction of screening (i.e. in the Netherlands 1989).

The progression of cancer is modelled using a semi-Markov process. After the onset of breast cancer a woman has preclinical pre-invasive disease. Preclinical pre-invasive disease can progress to preclinical invasive disease (progression from pre-invasive disease to invasive disease, not clinically manifest), clinical pre-invasive disease (if it becomes clinically manifest disease), or it may regress back to normal. The possibility of spontaneous regression has been shown in vitro, and there are case-reports on spontaneous

regression.(25, 26) In the situation with screening, preclinical disease can now become screen-detected (Figure 7).

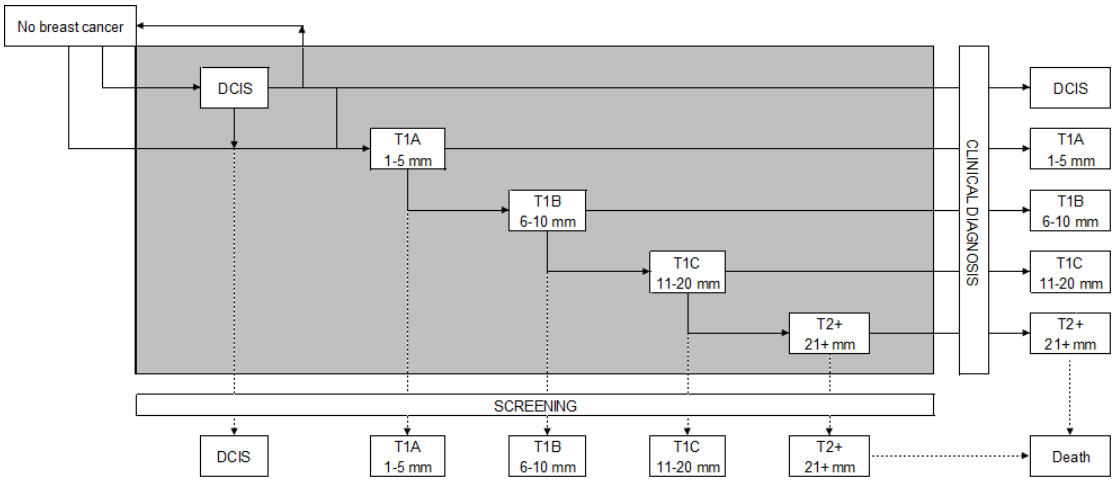


Figure 7. Progression of disease in the MISCAN model. DCIS: ductal carcinoma in situ.

The time between transitions cannot directly be observed. These durations, in the diagram in the gray area, are estimated based on stage distribution in a situation with screening.

The model is recalibrated regularly. Baseline onset is modelled on incidence rates prior to the introduction of screening. For the Dutch model an annual percentage change of 1.4% is modelled to account for increasing background incidence. In the Norwegian model an additional risk factor was introduced to account for the steep increase in breast cancer incidence. The sensitivity of the screening program is determined by the number of interval cancers within two years since the last screening examination. Stage specific breast cancer mortality and improvement of survival modeled using several international sources.(27, 28)

The model can calculate every event and every interval, for example number of breast cancers detected, number of (breast cancer) deaths, and number of invitations and screens, and stage distribution, all by age and calendar year. All output can be generated for a situation with screening, and for a situation without screening for the same population. With this output we can calculate the number of breast cancer deaths prevented by screening, the life years gained, overdiagnosis, etc.

With the output data we can calculate overdiagnosis is two ways. The model can directly provide output on women who have a screen-detected cancer which would never have been detected in the situation without screening, because her death would occur prior to the diagnosis of breast cancer in the situation without screening. Or we can calculate the number of breast cancer diagnoses in women of all ages in a situation without screening and subtract that from the number of breast cancer diagnoses in women of all ages in a situation with screening (Figure 6).

COST-EFFECTIVENESS ANALYSIS

A cost-effectiveness analysis is performed to determine if a health intervention is cost-effective. In a cost-effectiveness analysis the benefits of the intervention is compared to the costs of the intervention.(29) In the case of breast cancer screening the intervention is the screening program.

Benefits are measured in life years gained after adjusting for loss in quality of life (QALYs). The benefits of breast cancer screening are avoided breast cancer deaths, resulting in life years gained. These life years gained however are not all lived in good health. Diagnosis and treatment bear down on the quality of life. Utilities are measures to quantify the loss of quality of life. Utilities have a quantity and duration (Table 2). If an intervention reduces the quality of life by 15% for the duration of one half year the utility for that year is 0.5*0.85=0.43. Multiplying the life years gained by the utilities results in the QALYs gained.

To determine what is cost-effective, several measures can be used. The NICE guideline from 2013 advocates 20,000 GBP-30,000 GBP per QALY gained, which is 23,978 to 35,967 EUR (exchange rate July 14th 2016).(30) This threshold is gradually expanding. The World Health Organization uses a threshold of three times the gross domestic product (GDP) per capita to estimate cost-effectiveness.

Table 2. Durations and utilities for breast cancer screening and treatment, from Haes et al.(31)

Health stage	Duration	Utility
Screening	0.0192	0.9940
Diagnostic phase	0.0962	0.8950
Initial surgery	0.1667	0.1330
Initial radiotherapy	0.1667	0.1970
Initial chemotherapy	0.5000	0.2830
Initial hormonal therapy	2.0000	0.1800
Disease free 2m-1y mastectomy	0.8333	0.1560
Disease free 2m-1y breast conserving	0.8333	0.0860
Disease free >1y mastectomy †	1.0000	0.0530
Disease free >1y breast conserving †	1.0000	0.0400
Terminal illness	0.0833	0.7120
Palliative therapy + chemotherapy	0.3333	0.4690
Palliative therapy + radiotherapy	0.0833	0.4190
Palliative therapy + surgical therapy	0.0962	0.3830
Palliative therapy + hormonal therapy	1.1667	0.3370

RESEARCH QUESTIONS AND OUTLINE OF THIS THESIS:

The aim of the thesis is to evaluate overdiagnosis in breast cancer screening in well-established national screening programs. To achieve this we addressed the following research questions:

The detection of DCIS seems to be associated with a higher risk of overdiagnosis. What is the impact of the transition to digital mammography on the amount of DCIS diagnosed with screening (Chapter 2)?

Overdiagnosis is a well debated issue in breast cancer screening, but not in screening for cervical cancer screening. Is there overdiagnosis in cervical cancer screening, and how does this relate to the overdiagnosis rate in breast cancer screening (Chapter 3)?

What is the impact of DCIS on overdiagnosis estimates in breast cancer screening (Chapter 4)?

Is the Norwegian Breast Cancer Screening Program generating overdiagnosis, and at what level (Chapter 5)?

Is the NBCSP effective in reducing breast cancer mortality despite the occurrence of overdiagnosis (Chapter 6)?

REFERENCES

1. Bland KI, Copeland EM. The breast : comprehensive management of benign and malignant diseases. 4th ed. Philadelphia, PA: Saunders Elsevier; 2009.
2. Breast Cancer: American Cancer Society; 2015 [updated 10/06/2015. Available from: http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-breast-cancer-types.
3. Edge S BD, Comptom CC,Frits AG, Greene FL, Trotti A. AJCC Cancer Staging Manual. 7 ed: Springer-Verlag New York; 2010. XV, 648 p.
4. Guideline Breast Cancer Care: IKNL; [Available from: http://richtlijnendatabase.nl/en/richtlijn/breast_cancer/breast_cancer.html.
5. Cijfersoverkanker: Dutch Cancer Registry; 2016 [Available from: http://cijfersoverkanker.nl.
6. Anampa J, Makower D, Sparano JA. Progress in adjuvant chemotherapy for breast cancer: an overview. BMC Med. 2015;13:195.
7. 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(11):2205-40.
8. Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, et al. Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med. 2015;372(24):2353-8.
9. Otto SJ, Fracheboud J, Verbeek AL, Boer R, Reijerink-Verheij JC, Otten JD, et al. Mammography screening and risk of breast cancer death: a population-based case-control study. Cancer Epidemiol Biomarkers Prev. 2012;21(1):66-73.
10. 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(7):491-9.
11. Autier P, Boniol M, Gavin A, Vatten LJ. 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.
12. Otto SJ, Fracheboud J, Looman CW, Broeders MJ, Boer R, Hendriks JH, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. Lancet. 2003;361(9367):1411-7.
13. Smallenburg V, Duijm LE, den Heeten GJ, Groenewoud JH, Jansen FH, Fracheboud J, et al. Two-view versus single-view mammography at subsequent screening in a region of the Dutch breast screening programme. Eur J Radiol. 2012;81(9):2189-94.
14. de Gelder R, Heijnsdijk EA, Fracheboud J, Draisma G, de Koning HJ. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. Int J Cancer. 2015;137(1):165-72.
15. Zahl PH, Maehlen J. Overdiagnosis of breast cancer after 14 years of mammography screening. Tidsskr Nor Laegeforen. 2012;132(4):414-7.
16. Research-based evaluation of the Norwegian Breast Cancer Screening Program. The Research Council of Norway, May, 2015. Report No.
17. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol. 2010;49(5):725-36.

18. Hofvind S, Lee CI, Elmore JG. Stage-specific breast cancer incidence rates among participants and non-participants of a population-based mammographic screening program. Breast Cancer Res Treat. 2012;135(1):291-9.
19. Tornberg S, Kemetli L, Lynge E, Helene Olsen A, Hofvind S, Wang H, et al. Breast cancer incidence and mortality in the Nordic capitals, 1970-1998. Trends related to mammography screening programmes. Acta Oncol. 2006;45(5):528-35.
20. Moger T. Direct and indirect costs of the Norwegian Breast Cancer Screening Program. : University of Oslo; 2012(3) [Available from: http://www.med.uio.no/helsam/forskning/nettverk/hero/publikasjoner/skriftserie/2012/hero2012-3.pdf.
21. 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(1):111-21.
22. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am. 2004;42(5):793-806, v.
23. 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.
24. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985;20(1):79-93.
25. Burnside ES, Trentham-Dietz A, Kelcz F, Collins J. An Example of Breast Cancer Regression on Imaging. Radiology Case Reports. 2006;1(2):27-37.
26. Dehen R. Regression of Ductal Carcinoma In Situ After Treatment with Acupuncture. J Altern Complement Med. 2013.
27. de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst. 1995;87(16):1217-23.
28. Tabar L, Yen MF, Vitak B, Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet. 2003;361(9367):1405-10.
29. Habbema DF. A simulation approach to cost-effectiveness and cost benefit calculations of screening for early detection of disease. Eur j oper res. 1987;29(2):159-66.
30. Guide to the methods of technology appraisal 2013. National Institute for Health and Care Excellence; 2013.
31. de Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van Der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. Int J Cancer. 1991;49(4):538-44.

CHAPTER 2:

NATION-WIDE DATA ON SCREENING PERFORMANCE DURING THE TRANSITION TO DIGITAL MAMMOGRAPHY. OBSERVATIONS IN 6 MILLION SCREENS.

Paula van Luijt,

Jacques Fracheboud

Eveline Heijnsdijk

Gerard den Heeten

Harry de Koning

Eur J Cancer. 2013;49(16):3517-25.

ABSTRACT

PURPOSE: To critically evaluate and confirm previous results regarding the diagnostic accuracy of digital mammography screening (DM), compared to screen-film mammography (SFM) in the whole Dutch screening program, in the period 2004-2010, during which a full transition from SFM to DM was made.

MATERIALS AND METHODS: 1.5 million DM and 4.6 million SFM were read in the Dutch national breast-cancer screening program in the period 2004-2010. We evaluated recall rate, detection rate, positive predictive value and tumor-size distribution for younger and older women, for first time participants and women having a timely subsequent screen. We compared DM screens read by radiologists reading DM and SFM (DM-group) to SFM screens read by these radiologists (SFM-group) and to SFM screens read by radiologists reading only SFM (SFMonly-group).

RESULTS: Recall rate was 2.0 %(95% C.I.: 2.0; 2.1) in the DM-group, compared to 1.6% (95% C.I.: 1.6; 1.6) in the SFM-group and 1.6% (95% C.I.: 1.5; 1.6) in the SFMonly-group. The overall detection rates were 5.9/1,000 screens (95% C.I.: 5.7; 6.0) in the DM-group, 5.1/1,000 screens (95% C.I.: 5.0; 5.2) in the SFM-group and 5.0/1,000 screens (95% C.I.: 5.0; 5.1) in the SFMonly-group. Detection rate rose most markedly in younger women (age 49-54) from 4.0 / 1,000 screens to 5.1 / 1,000 screens (p-value <0.001). PPV in DM rose from 18.4% (95% C.I.: 14.6; 23.1) in 2004 to 32.5% (95% C.I.: 31.7; 33.2) in 2010. Detection rate rose in SFM-group from 5.0/ 1,000 screens (95% C.I.: 4.7; 5.3) in 2004 to 5.5/ 1,000 screens (95% C.I.: 5.2; 5.7) in2010. Detection rate in DM-group rose mostly due to DCIS detection especially in younger women/ first screens. The proportion of T1a tumors was significantly higher in DM-group; otherwise size distribution did not change significantly for invasive carcinoma. Recall rates were variable between different screening regions.

CONCLUSION: In accordance to previous, smaller, studies, we can confirm that DM has a higher detection rate compared to SFM, at the cost of a higher recall rate and lower PPV. More DCIS and a higher fraction of very small tumors were detected with DM, which has positive consequences for the stage shift as a result of mass screening.

INTRODUCTION

The independent UK Panel on Breast Cancer Screening recently published their review[1]. They state that in light of the current debate on breast cancer screening the benefits outweigh the harms. This panel based their decisions on data from screen-film mammography (SFM), some of which from decades ago. The consequences of a transition of the program to digital mammography (DM) were not yet fully elucidated.

The Dutch screening program was implemented in 1990. All targeted women aged 50-75 receive an invitation for screening mammography, free of charge, every other year. The attendance rate is approximately 80%. A feasibility study was carried out between 2003 and 2007 to ascertain whether it was feasible to convert to digital mammography in the screening program[2]. Since June 2010 all mammography screening in the Netherlands is performed digitally. In the recent past five studies were performed in the Netherlands, based on data of the feasibility study or from a single region[2-5]. These studies all found a higher recall rate, higher detection rate and lower PPV. Decision making in service screening in Europe is based mainly on population based data. Sometimes in depth information can only be manually extracted because it is not readily available in national data. The disadvantage of this is a lack of power. Since we observe considerable regional differences we wondered how a very large, national data set would relate to national and international communications on differences between SFM and DM.

All over Europe and elsewhere studies were conducted on the impact of a transition to digital mammography screening[2-12]. All studies found that the diagnostic precision of DM was at least similar to that of SFM, however some found a higher recall rate and lower positive predictive value (PPV)[4-6], where others found a similar or higher PPV[2,3,7-11]. No clear cut results on invasive stages/ DCIS were published. The differences in published outcomes may be explained by sample size and study design.

We retrospectively used the annual monitoring parameters in more than 1.5 million DM and 4.6 million SFM to assess performance during the transition from SFM to DM. As our study covers all screening examinations performed within the nation-wide breast cancer screening program and therefore has a larger external validity, also it enables us to compare differences between regions.

MATERIALS AND METHODS

CASE SELECTION

All women aged 50 to 75 (corresponding with ages 49-74 at the beginning of a calendar year) receive a personal invitation biennially for mammography screening. By participating in the program, women automatically consent to the use of their data to evaluate and improve the program. Information about the use of data is provided in a flyer accompanying the invitation letter. If a woman does not want the screening organization to use her data for this purpose, she can return the signed corresponding form to the screening organization. Only a minor fraction (0.01%) used this possibility[13].

We retrospectively evaluated the aggregated data over the years 2004-2010 from the regional screening organizations, which is delivered to us annually for the purpose of monitoring the nation-wide program. We had information on age of the women invited and attending, screening round and screening results, whether the examinations were digital or screen-film, and if radiologists were reading DM (always in combination with SFM at some point in time), SFM (in this case in combination with DM at some point in time), or only SFM (a group that diminished to zero in 2010).

In 2007 the roll out of digital screening was initiated nationwide, one screening unit at a time. Because several screening units are attached to one radiologists' group (so-called reading unit), this gradual roll out resulted in reading units reading both DM and SFM at a given time.

This gave us the opportunity to analyze our data by type of reading unit that evaluated the screens;

1. DM read by a reading unit reading both SFM and DM, we will refer to this group as DM-group.
2. SFM read by a reading unit reading both SFM and DM, we will refer to this group as SFM-group.
3. SFM read by a reading unit reading only SFM, we will refer to this group as SFMonly-group.

We performed subgroup analyses on women attending their first screening examinations aged 49-51, on women attending subsequent screens performed within the appropriate screening interval of 2.5 years (we will refer to these as timely subsequent screens) aged 50-74, on younger women aged 49-54, and on older women aged 55-74 for first and subsequent screens together.

Our main outcome measures were recall rate (the number of women recalled for clinical assessment, based on suspicion of malignancy and inconclusive results per 1,000 women screened), detection rate (the number of women diagnosed with breast cancer as a result of screening per 1,000 women screened), PPV (the percentage of recalled women diagnosed with breast cancer), and stage distribution of screen detected breast cancers (based on size and invasion).Because detection rate overall increased, stage distribution was expressed by the proportion of all detected tumors to indicate differences in stage shift, rather than increased detection.

The data contained information on the reading unit (reading SFM, DM or both), on the number of women initially or subsequently screened, on the interval period between screens, and on the number of recalls and screen-detected breast cancers. From the 6,370,556 screens performed between 2004 and 2010, we excluded women older than 74, because they represent very small numbers (n= 7,548), women who had their first screen at an age older than 51 years (n= 90,665), and women with a subsequent screen after a screening interval of more than 2.5 years (n= 264,761). The latter two groups were expected to have a higher detection rate, dependent on age-specific incidence and the average screening interval.[13] After exclusion 6,007,582 screens were left for analysis.

SCREENING RADIOLOGISTS

A Dutch screening radiologist has to be certified for screening in the national service screening program. This can be achieved by participating in an eight days training in the National Training and Reference Center (NETC) in Nijmegen. The mean reading volume is 13,000 screens per year with a range from 3,000 to 60,000. Increasing numbers of screening organisations are now routinely performing two view mammographies. In 2010 93% of all screens were performed with two view mammography.

DATA ANALYSIS

We retrospectively analyzed the aggregated data for the DM and SFM screens for the following outcome measures: recall rate, detection rate, positive predictive value, and tumor size distribution of screen-detected carcinomas. Each screening examination is one event, regardless of the number of mammography views, lesions or malignancies.

During the study period 2004-2010, we compared outcome measures and trends between the SFM-group, DM-group, and SFMonly-group and between first screens, timely subsequent screens, and first screens and timely subsequent screens together. Additionally we stratified results by age-group, 49-54 and 55-74.

Because the detection of DCIS rose markedly we decided to analyse the proportion of disease stages in invasive breast cancer separately from the DCIS. The proportion of DCIS is of the total number of screen detected breast cancer, the proportions of T1a, T1b, T1c and T2+ are of the sum of all invasive cancers (explicitely not including DCIS).

SPREAD OF OUTCOMES BY REGION

We also looked at the spread of outcomes between different regions. During the study period, the Netherlands had nine different screening regions. They were allocated letters A through I. Each screening region contained a variable number of reading units. The results by region were divided in DM-group and SFM-group, for the SFMonly- group we did not have data per region, and by women aged 49-54 and women aged 55-74.

STATISTICAL ANALYSIS

All rates have been age-adjusted by direct standardization using the Dutch female population in 2000 as a reference population. We used linear regression to test for differences in proportions between DM and SFM and we stratified by age. P values less than 0.05 were considered significant. All regression analyses were done using R.

RESULTS

In total 6,007,582 screens were included for analysis. Of these 1,452,508 (24%) were DM, 1,460,344 (24%) SFM in the SFM-group, and 3,094,730 (52%) SFM in the SFMonly-group (Table 1). There was no difference between the groups regarding age and the distribution of first and subsequent screens. In 2010 there were no more reading units reading only SFM.

Table 1. Descriptive statistics. The percentage indicated in screening round is the percentage of all screens in that group, the percentage indicated in the number of screens is the percentage of a group of all screens in that year. Timely screen= a subsequent screen within 2.5 years since the previous screen. DM= digital mammography, SFM= screen-film mammography. DM-group= DM-screens read by radiologists reading both SFM and DM, SFM-group= SFM-screens read by radiologists reading both SFM and DM, SFMonly-group= SFM screens read by radiologists reading only SFM.

	DM N= 1,452,508		SFM N= 1,460,344		SFMonly N= 3,094,730	
Mean age	μ 59.69		μ 59.62		μ 59.58	
	N	%	N	%	N	%
Screening round						
First screen	164,652	11	165,225	11	351,892	11
Timely subsequent screen	1,287,856	89	1,295,119	89	2,742,838	89
Number of screens						
2004	9,342	1	177,180	21	643,998	78
2005	33,087	4	161,015	19	643,541	77
2006	39,494	5	153,337	18	640,800	77
2007	62,744	7	175,429	20	621,628	72
2008	87,055	10	364,710	42	419,069	48
2009	364,041	42	375,360	43	125,694	15
2010	856,745	94	53,313	6	n.a.	n.a.
2004-2010	1,452,508	24	1,460,344	24	3,094,730	52

OVERALL

With regard to the overall results of the three different groups of reading units, across the years the DM-group had a significantly higher recall rate of 2.0% (95% C.I. 2.0; 2.1) than both the SFM-group with 1.6% (95% C.I. 1.6; 1.6), and the SFMonly-group with 1.6% (95% C.I. 1.5; 1.6) (Table 2). This was also true for detection rate (5.9/1,000 screens in DM-group (95% C.I. 5.7; 6.0), vs. 5.1 (95% C.I 5.0; 5.2), and 5.0 (95% C.I. 4.9; 5.1) in SFM-group and SFMonly-group respectively). PPV was significantly lower in DM-group at 31.2% (95% C.I. 30.6; 31.7) compared to 34.4% (95% C.I. 33.8; 35.0), and 34.2% (95% C.I. 33.7; 34.6) in SFM-group and SFMonly-group respectively.

TRENDS

In the DM-group the recall rate dropped rapidly from 3.41,000% (95% C.I. 3.1; 3.8) in 2004 to 2.01,000% (95% C.I. 2.0; 2.0) in 2010, and the detection rate dropped from 6.5/ 1,000 screens (95% C.I. 5.1; 8.4) in 2004 to 5.9/ 1,000 screens (95% C.I. 5.7; 6.1) in 2010 after an initial increase in the pilot, stabilizing at a lower level (Figure 1a and 1b), the resulting PPV rose accordingly from 18.5% (95% C.I. 14.6; 23.1) in 2004 to 32.5% (95% C.I 31.7; 33.2) in 2010 and also stabilized, but at a higher level (Figure 1c). In the SFM-group recall rate and detection rate rose up to and including 2009 (recall rate went from 1.5 /1,000 (95% C.I.: 1.4; 1.6) in 2004 to 1.7 /1,000 (95% C.I. : 1.7; 1.8), detection rate went from 5.0 /1,000 (95% C.I.: 4.7; 5.3) to 5.5 (95% C.I.: 5.3; 5.8)), resulting in an overall unaltered PPV (35.1% (95% C.I.: 33.9; 36.3) in 2009). In 2010, all three parameters dropped, probably as a result of smaller numbers of screening examinations (see Table 1). Finally, in the SFMonly-group recall rate increased, from 1.3% (95% C.I. 1.3; 1.3) in 2004 to 1.8% (95% C.I. 1.8; 1.9) in 2009, but detection rate did not improve, thus PPV decreased over time (from 37.9% (95% C.I.: 36.8; 38.9) in 2004 to 29.4% (95% C.I.: 27.6; 31.4) in 2009).

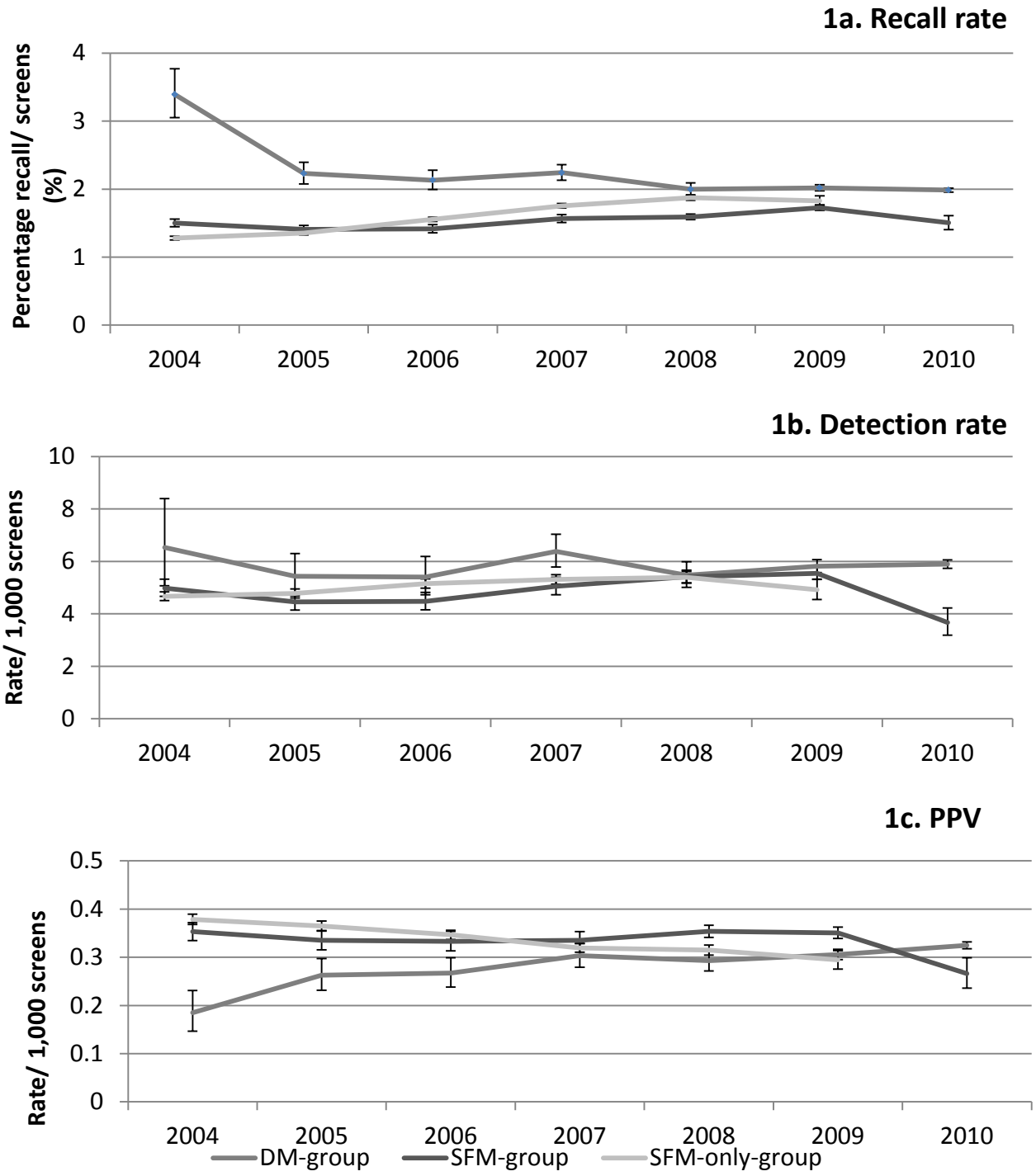


Figure 1. Recall rate per 1,000 screens (1a), detection rate per 1,000 screens (1b) and positive predictive value at screening (PPV, %) (1c), development over time. DM= digital mammography, SFM= screen-film mammography. DM-group= DM-screens read by radiologists reading both SFM and DM, SFM-group= SFM-screens read by radiologists reading both SFM and DM, SFMonly-group= SFM screens read by radiologists reading only FSM.

SUBGROUP ANALYSIS BY AGE AND BY SCREENING ROUND

We stratified the analysis by age (younger women, 49-54 years, and older women, 55-74 years), and by screening round (first screens in women aged 49-51 years, and timely subsequent screens in all ages).

RECALL RATE

In all subgroup analyses recall rate was significantly higher in the DM-group than in the other groups, with the exception of timely subsequent screens (Table 2). The biggest difference (p-value <0.001) was found in first screens with a recall rate for DM-group of 4.6% compared to 3.3% in SFM-group and 3.2% in SFMonly-group (Table 2).

DETECTION RATE

Detection rate was significantly higher (p-value< 0.001) for DM-group in all subgroup analyzes (Table 2). This was the most pronounced in younger women (5.1/1,000 vs. 4.0/1,000 in SFM-group and 4.1/1,000 in SFMonly-group), and in first screens (7.0/1,000 in DM-group vs. 5.5/1,000 in SFM-group and 5.6/1,000 in SFMonly-group).

PPV

PPV was significantly lower in DM-group than in the other groups in all subgroup analyzes, except in younger women (Table 2). This difference (p-value < 0.001) was largest in older women (35.7% in DM-group, vs. 40.1% in SFM-group, and 39.4% in SFMonly-group), and in subsequent screens (p-value < 0.001) (30.6% in DM-group, vs. 33.5% in SFM-group, and 33.4% in SFMonly-group).

Table 2. Overall results and results by subgroup: women aged 49-54, women aged 55-74, first screens and timely subsequent screens. P-values less than 0.05 are considered significant and indicated in bold. DM= digital mammography screen, SFM= screen-film mammography screen, DM-group= DM-screens read by radiologists reading both SFM and DM, SFM-group= SFM-screens read by radiologists reading both SFM and DM, SFMonly-group= SFM screens read by radiologists reading only SFM. PPV= positive predictive value.

		Screens	Recalls	Recall rate (%)	P-value	Cancers	Detection rate/ 1,000 screens	P-value	PPV	P-value
Overall	DM-group	1,452,508	29,363	2.0		8,474	5.9		31.2%	
	SFM-group	1,460,344	22,799	1.6	<0.001	7,382	5.1	<0.001	34.4%	<0.001
	SFMonly-group	3,094,730	47,683	1.6	<0.001	15,407	5.0	<0.001	34.2%	<0.001
49-54 years (first screens 49-51 years, timely subsequent screens 50-74 years)	DM-group	429,692	11,973	2.7		2,247	5.1		21.4%	
	SFM-group	434,064	8,882	2.0	<0.001	1,773	4.0	<0.001	22.1%	0.23
	SFMonly-group	932,129	18,369	1.9	<0.001	3,835	4.1	<0.001	23.0%	<0.001
55-74 years (timely subsequent screens)	DM-group	1,022,816	17,390	1.7		6,227	6.2		35.7%	
	SFM-group	1,026,280	13,917	1.4	<0.001	5,609	5.6	<0.001	40.1%	<0.001
	SFMonly-group	2,162,601	29,314	1.4	<0.001	11,572	5.5	<0.001	39.4%	<0.001
First screen (49-51 years)	DM-group	164,652	7,422	4.6		1,103	7.0		15.2%	
	SFM-group	165,225	5,359	3.3	<0.001	880	5.5	<0.001	16.5%	0.05
	SFMonly-group	351,892	10,773	3.2	<0.001	1,862	5.6	<0.001	17.6%	<0.001
Timely subsequent screen (50-74 years)	DM-group	1,287,856	21,941	1.6		7,371	5.3		30.6%	
	SFM-group	1,295,119	17,440	1.3	<0.001	6,502	4.6	<0.001	33.5%	<0.001
	SFMonly-group	2,742,838	36,910	1.7	<0.001	13,545	4.6	<0.001	33.4%	<0.001

STAGE DISTRIBUTION

In both subgroup analyses significantly more DCIS (p-value < 0.001) were detected in DM-group (Figure 2): in first screens 24.5% vs. 19.7% in SFM-group, and 18.2% in SFMonly-group21.5% in DM-group vs. 15.6% in SFM-group, and 14.8% in SFMonly-group; in younger women 25.9% in DM-group vs. 19.5% in SFM-group and 19.1% in SFMonly-group; and in older women 19.3% in DM-group vs. 14.5% in SFM-group, and 13.2% in SFMonly-group.

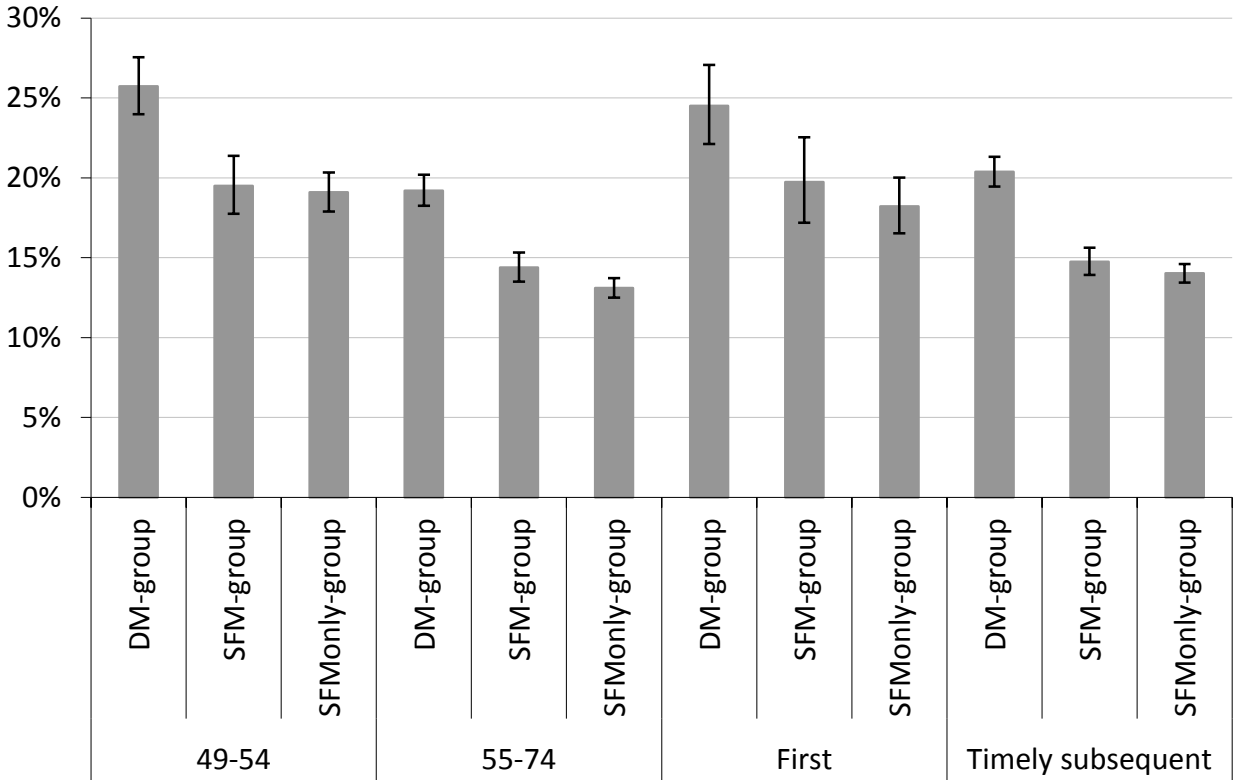


Figure 2. Proportion of DCIS (%, with 95% confidence interval) of all screen-detected cancers by subgroup for younger and older women and first and timely subsequent screen. DM= digital mammography screen, SFM= screen-film mammography screen, DM-group= DM-screens read by radiologists reading both SFM and DM, SFM-group= SFM-screens read by radiologists reading both SFM and DM, SFMonly-group= SFM screens read by radiologists reading only SFM.

The analysis of the invasive cancers was done separately from DCIS, to avoid the large increase in DCIS to affect the proportions of invasive carcinomas. T1a was found significantly more frequently in DM-group in both age-groups (7.2% of invasive tumors in younger women vs. 5.3% in SFM-group (p-value= 0.02), and 5.6% in SFMonly-group (p-value= 0.04), and 6.9% in older women vs. 4.7% in SFM-group (p-value < 0.001), and 5.0% in SFMonly-group(p-value < 0.001)), as well as in subsequent screens (4.4% in DM, vs 6.9% in SFM and 5.7% in SFMonly, p-value < 0.001) (Table 3).

Table 3. Tumor size distribution of invasive tumors by subgroup, including unclassified tumors (not shown). P-values less than 0.05 are considered significant. N= number of cases, %= age adjusted proportion of that size of all detected breast cancers (without the tumors with unknown stage and size), 95% C.I.= 95% confidence interval, DM= digital mammography, SFM= screen-film mammography. DM-group= DM-screens read by radiologists reading both SFM and DM, SFM-group= SFM-screens read by radiologists reading both SFM and DM, SFMonly-group= SFM screens read by radiologists reading only SFM. T1a= tumor with a diameter up to 0.5 cm, T1b= tumor with a diameter from 0.6 cm up to 1.0 cm, T1c= tumor with a diameter from 1.1 cm up to 2.0 cm, T2+= any tumor with a diameter larger than 2.0 cm.

		Women aged 49-54			Women aged 55-74			First screen			Timely subsequent screen		
		N	%	P-value	N	%	P-value	N	%	P-value	N	%	P-value
T1a	DM-group	120	7.2%		346	6.9%		60	7.1%		406	4.4%	
	SFM-group	74	5.3%	0.02	224	4.7%	<0.001	36	4.1%	0.0126	262	6.9%	<0.001
	SFMonly-group	174	5.6%	0.04	506	5.0%	<0.001	81	6.4%	0.5184	599	5.7%	<0.001
T1b	DM-group	302	17.9%		1,272	25.0%		133	16.0%		1,441	21.9%	
	SFM-group	286	20.2%	0.08	1,109	22.8%	0.014	129	18.4%	0.215	1,266	22.6%	0.38
	SFMonly-group	582	18.7%	0.56	2,278	22.4%	<0.001	269	17.7%	0.3	2,591	19.8%	<0.001
T1c	DM-group	827	50.0%		2,417	47.8%		393	48.0%		2,851	46.8%	
	SFM-group	708	50.0%	0.87	2,441	50.8%	0.004	342	48.1%	0.97	2,807	47.1%	0.67
	SFMonly-group	1,526	49.4%	0.6	5,263	52.2%	<0.001	727	47.4%	0.78	6,062	49.6%	<0.001
T2+	DM-group	362	21.8%		866	17.5%		191	25.2%		1,037	16.8%	
	SFM-group	308	21.3%	0.64	916	19.0%	0.04	171	23.5%	0.44	1,053	17.9%	0.13
	SFMonly-group	703	22.6%	0.53	1,829	18.4%	0.23	364	23.8%	0.45	2,168	18.0%	0.04

In older women, DM detected significantly more T1b tumors (25.0%) than both the SFM-group (22.8%, p-value < 0.014) and SFMonly-group (22.4% p-value < 0.001).

In younger women, and by stratifying by screening round, no significant differences in the proportion of T1b, T1c and T2+ were found.

As a consequence significantly less T1c and T2+ tumors were found in DM-group in older women: 47.8% T1c tumors in DM-groupvs. in SFM-group (50.8%, p-value = 0.004), and SFMonly-group (52.2%, p-value < 0.001), and 17.5% T2+ vs. 19.0% in SFM-group (p-value = 0.04). In the timely subsequent screens the difference in proportion of T2+ tumors was significantly lower for DM, 16.8% in DM-group, vs. 18.0% in SFMonly-group (p-value = 0.04).

The number of screen detected breast cancers with a positive node status did not differ significantly between the groups (data not shown).

VARIATION BETWEEN REGIONS

We evaluated the results by screening region, the largest spread is found in recall rate, with less variation in detection rate, resulting in varying PPV (Table 4). The spread is given in overall numbers for the entire study period. When looking at the range of variation over the years of the study period we did not find evidence for convergence (data not shown).

Table 4. Point estimates and range of variation in regions in recall rate, detection rate and PPV between regions. DM= digital mammography, SFM= screen-film mammography. DM-group= DM-screens read by radiologists reading both SFM and DM, SFM-group= SFM-screens read by radiologists reading both SFM and DM. PPV= positive predictive value.

	DM-group		FSM-group	
	National	Range of variation	National	Range of variation
Women aged 49-54				
Recall rate (%)	2.7	1.7-3.3	2.0	1.3-3.0
Detection rate (%)	5.1	4.4-6.2	4.0	3.4-4.8
PPV (%)	21.4	16.8-29.9	22.1	15.0-30.8
Women aged 55-74				
Recall rate (%)	1.7	1.2-2.0	1.4	1.1-1.6
Detection rate (%)	6.2	5.5-6.7	5.6	5.0-6.4
PPV (%)	35.7	30.5-47.5	40.1	30.8-54.0

DISCUSSION

In this study, we were able to confirm the higher detection rate for DM at the cost of a higher recall rate, and a slightly lower positive predictive value of recall. But most importantly we found a significant difference in the detection of DCIS and T1a tumors.

After an initial increase, the recall rate declined in the years following first introduction of DM. This is in line with the objective of the Netherlands Expert and Training Centre for Breast cancer screening (NETCB), advising to aim for an overall recall rate of approximately 20 per 1,000 screens. The NETCB offered one day specialist training courses for the reading radiologists on reading DM to obtain this objective.

The recall rate in the Dutch program has stabilized at 20/1,000 screens. In an international perspective, this is still relatively low[14]. Incidence rates of invasive breast cancer in the Netherlands (both screen detected and clinically detected) remained stable throughout the study period (270-400/100,000 women, dependent on age), incidence rates of DCIS increased steadily with the expansion of DM within the country (from 40-45/100,000 women to 55-68/100,000 women, dependent on age)[15].

When we place the recall rate in the perspective of long term performance data of the Dutch program we see that there has been a pre-existing trend towards higher recall rates since the second half of the 1990s. This trend does not appear to be strongly affected by the introduction of digital screening[13]. We confirmed the results of the earlier studies with regards to recall rate, detection rate and PPV[2,3,5,7]. We did not have the data to determine the grade of DCIS detected with DM. However we found a significantly higher detection rate of small invasive tumors[7].

In the perspective of the European literature on the subject, with 152,515 DM being the largest reported study, we can confirm previous results with regard to diagnostic precision. We found that PPV is slightly lower in DM, but steadily increases over time. This must be due to a learning curve with better understanding of the findings on DM. Compared to different countries the recall rate is relatively low in the Netherlands, with a high PPV (Table 5)[6,8-11,16-18].

Table 5. Summary of some European research on performance of digital mammography, compared to screen-film mammography. SFM= the number of film screen examinations analyzed, DM= the number of digital screen examinations analyzed. RR=recall rate, DR= detection rate, PPV= positive predictive value, FP= false positive rate.

Author	Reference	Year	Country	Study Population		RR (%)	DR (%)	PPV (%)	FP (%)
Ongeval	16	2010	Belgium	DM	11,355	2.1	0.59	34.9	
				SFM	23,325	1.58	0.64	30.76	
Domingo	12	2011	Spain	DM	71,647	6.1	0.43	7.0	
				SFM	171,191	8.0	0.45	6.0	
Sala	17	2009	Spain	DM	6,074	4.2	0.4		3.8
				SFM	12,958	5.5	0.4		5.1
Vinnicombe	10	2009	Great Britain	DM	88,478	4.79	0.68	14.3	
				SFM	31,720	4.43	0.65	14.6	
Hambly	9	2009	Republic of Ireland	DM	35,204	4.0	6.3	15.7	
				SFM	153,619	3.1	5.2	16.7	
del Turco	11	2007	Italy	DM	14,385	4.56	0.72	15.9	
				SFM	14,385	3.96	0.58	14.7	
Skaane	18	2004	Norway	DM	6,997	3.8	0.59	21.6	
				SFM	17,911	2.5	0.41	22.1	
Vigeland	8	2007	Norway	DM	18,239	4.09	0.77	18.5	
				SFM	324,763	4.16	0.65	15.1	
Our results		2013	The Netherlands	DM	1,452,508	2.0	0.59	31.2	
				SFM	1,460,344	1.6	0.51	34.4	

Despite the fact that the radiologists in SFM-group and SFMonly-group did not change technique, there is an obviously increasing trend in recall rate and detection rate. This was in part intentional policy communicated during audits and onsite visits by the NETCB, as a result of the optimization study carried out in 2004 and perhaps also influenced by the recall rates in DM[2].

The sharpest increase was found in the proportion of DCIS detected. This immediately triggers the concern of overdiagnosis, as often raised by those opposing screening. Since this issue can only be elucidated by very long term evaluations of mortality reduction, at this moment only modeling studies can be used to predict the impact of DM on overdiagnosis rate. The clinical consequences of higher DCIS detection rates have been explored by de Gelder et al. in 2011[19]. She used the data of the feasibility study from 2004 to 2006 and a statistical model to predict the mortality reduction and overdiagnosis rate. The current increase in DCIS is still in line with the data used in their paper (1.2/1,000 screens) and does not alter their conclusion that increased detection of DCIS by DM reduces breast cancer mortality by a further 4.4% at a 21% increased overdiagnosis rate.

The increase in the amount of detected DCIS is stronger in younger women and in first screens (aged 50-51). This confirms the previous result of Pisano et al. in the US that the detection rate is more strongly affected by DM in younger women[12]. In oncology DCIS is still considered a serious condition and should be treated. [20]

The increase in the number of DCIS detected is rather steep. Possibly this is due to a first pass effect. This means that outcomes will be most strongly affected directly after the introduction of a new technology or method, comparable to a prevalence screen. The number of DCIS might stabilize at a lower level in the upcoming years.

As a result of higher DCIS detection rates, we expect the detection rates of more advanced tumors to decline in the upcoming years, although we do expect part of the DCIS to represent overdiagnosis[19,21]. A recent study of the type of DCIS in the pilot phase showed no shift towards low grade DCIS, but significantly more high grade DCIS in subsequent screens[7].

We also found significantly more T1a tumors in DM-group. Detection of invasive breast cancer in an early stage may be beneficiary for the results of screening on mortality reduction, dependent on the grade of the tumors detected. We have no information on grade, but Nederend et al. found that DM finds more low and intermediate grade tumors[5]. The effect of an increased detection rate on interval cancer rate needs to be awaited.

A wide range of variance exists when looking at regions separately. All regions perform within the internationally set standards. We found no evidence for convergence over time; this is in line with our expectations as multiple parameters vary (reading radiologists, changes in equipment etc.).

With the introduction of DM many screening organizations were stimulated to perform two view examinations, not only at initial screen, but also at regular follow up screen. In 2010 93% of all participants were examined using a two view examination. This change in policy may also have influenced the performance rates of the screening program[22].

In summary, we can confirm earlier results on DM screening in terms of diagnostic accuracy and can add that DM detects a significantly higher amount of DCIS and small invasive tumors. The effects of this additional stage shift on mortality reduction will have to be awaited. The performance of the Dutch screening program in international perspective is good, with a low recall rate and a high detection rate.

REFERENCES

1. Independent UKPoBCS. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; **380**(9855): 1778-86.
2. Bluekens AM, Karssemeijer N, Beijerinck D, et al. Consequences of digital mammography in population-based breast cancer screening: initial changes and long-term impact on referral rates. *Eur Radiol* 2010; **20**(9): 2067-73.
3. Karssemeijer N, Bluekens AM, Beijerinck D, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology* 2009; **253**(2): 353-8.
4. Timmers JM, den Heeten GJ, Adang EM, Otten JD, Verbeek AL, Broeders MJ. Dutch digital breast cancer screening: implications for breast cancer care. *Eur J Public Health* 2012; **22**(6): 925-9.
5. Nederend J, Duijm LE, Louwman MW, Groenewoud JH, Donkers-van Rossum AB, Voogd AC. Impact of transition from analog screening mammography to digital screening mammography on screening outcome in The Netherlands: a population-based study. *Ann Oncol* 2012; **23**(12): 3098-103.
6. Vigeland E, Klaasen H, Klingen TA, Hofvind S, Skaane P. Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study. *Eur Radiol* 2008; **18**(1): 183-91.
7. Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. *Radiology* 2012; **265**(3): 707-14.
8. Hambly NM, McNicholas MM, Phelan N, Hargaden GC, O'Doherty A, Flanagan FL. Comparison of digital mammography and screen-film mammography in breast cancer screening: a review in the Irish breast screening program. *AJR Am J Roentgenol* 2009; **193**(4): 1010-8.
9. Vinnicombe S, Pinto Pereira SM, McCormack VA, Shiel S, Perry N, Dos Santos Silva IM. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. *Radiology* 2009; **251**(2): 347-58.
10. Del Turco MR, Mantellini P, Ciatto S, et al. Full-field digital versus screen-film mammography: comparative accuracy in concurrent screening cohorts. *AJR Am J Roentgenol* 2007; **189**(4): 860-6.
11. Domingo L, Romero A, Belvis F, et al. Differences in radiological patterns, tumour characteristics and diagnostic precision between digital mammography and screen-film mammography in four breast cancer screening programmes in Spain. *Eur Radiol* 2011; **21**(9): 2020-8.
12. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. *Radiology* 2008; **246**(2): 376-83.
13. NETB. National evaluation of breast cancer screening in the Netherlands 1990-2007. Twelfth evaluation report., 2009.
14. Otten JD, Karssemeijer N, Hendriks JH, et al. Effect of recall rate on earlier screen detection of breast cancers based on the Dutch performance indicators. *J Natl Cancer Inst* 2005; **97**(10): 748-54.

15. Cijfersoverkanker. 2016. http://cijfersoverkanker.nl2016).
16. Van Ongeval C, Van Steen A, Vande Putte G, et al. Does digital mammography in a decentralized breast cancer screening program lead to screening performance parameters comparable with film-screen mammography? *Eur Radiol* 2010; **20**(10): 2307-14.
17. Sala M, Comas M, Macia F, Martinez J, Casamitjana M, Castells X. Implementation of digital mammography in a population-based breast cancer screening program: effect of screening round on recall rate and cancer detection. *Radiology* 2009; **252**(1): 31-9.
18. Skaane P, Skjennald A, Young K, et al. Follow-up and final results of the Oslo I Study comparing screen-film mammography and full-field digital mammography with soft-copy reading. *Acta Radiol* 2005; **46**(7): 679-89.
19. de Gelder R, Fracheboud J, Heijnsdijk EA, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med* 2011; **53**(3): 134-40.
20. Bijker N, Donker M, Wesseling J, den Heeten GJ, Rutgers EJ. Is DCIS breast cancer, and how do I treat it? *Current treatment options in oncology* 2013; **14**(1): 75-87.
21. Verbeek AL, Broeders MJ, Otto SJ, et al. [Effects of the population screening into breast cancer]

Effecten van het bevolkingsonderzoek naar borstkanker. *Ned Tijdschr Geneeskd* 2013; **157**(10): A5218.
22. Smallenburg V, Duijm LE, den Heeten GJ, et al. Two-view versus single-view mammography at subsequent screening in a region of the Dutch breast screening programme. *Eur J Radiol* 2012; **81**(9): 2189-94.

CHAPTER 3:

THE ROLE OF PRE-INVASIVE DISEASE IN OVERDIAGNOSIS: A MICROSIMULATION STUDY COMPARING MASS SCREENING FOR BREAST CANCER AND CERVICAL CANCER.

Paula van Luijt

Kirsten Rozemeijer

Steffie Naber

Eveline Heijnsdijk

Joost van Rosmalen,

Marjolein van Ballegooijen

Harry de Koning.

J Med Screen. 2016, april 11. pii: 0969141316629505. [Epub ahead of print]

ABSTRACT

BACKGROUND: Early detection of cancer prevents cancer deaths if an effective treatment is available for the early stage at detection. A drawback of mass screening is overdiagnosis. The potential harm of overdiagnosis depends on its frequency and the consequences of diagnosis and treatment. There is much debate on the topic of overdiagnosis in screening for breast cancer, but less so on overdiagnosis in screening for cervical cancer.

MATERIALS AND METHODS: We estimated overdiagnosis rates by microsimulation for breast cancer screening and for cervical cancer screening, using a cohort run of women born in 1982 with lifelong follow-up. Overdiagnosis estimates were made analogous to two definitions formed by the UK 2012 breast screening review. Pre-invasive disease was included in both definitions.

RESULTS: Screening prevented 921 cervical cancers (-55%); and 378 cervical cancer deaths (-59%) and 169 (-1.3%) breast cancer cases and 970 breast cancer deaths (-21%). Cervical cancer overdiagnosis rate was 74.8%, when including pre-invasive disease. Breast cancer overdiagnosis rate was estimated at 2.5%, when including pre-invasive disease. For women of all ages in breast cancer screening, an excess of 207 diagnoses/100,000 women was found with screening, compared to an excess of 3,999 diagnoses/100,000 women in cervical cancer screening.

CONCLUSION: For breast cancer, the frequency of overdiagnosis in screening is relatively low, but consequences are evident. For cervical cancer, the frequency of overdiagnosis in screening is high, because of detection of pre-invasive disease, but the consequences per case are relatively small due to less invasive treatment. This illustrates that it is necessary to present overdiagnosis in relation to disease stage and consequences.

INTRODUCTION

The purpose of cancer screening is to prevent cancer death by detecting a cancerous lesion early, and for some cancers pre-cancerous lesions, when treatment is still a viable option and more effective, or cancer may be prevented altogether.(1) Screening advances the diagnosis of disease to an earlier age, resulting in a higher incidence just after the initiation of screening. After the upper age limit of screening, the incidence rate will drop.(2)

Breast cancer screening detects invasive breast cancer and ductal carcinoma in situ (DCIS), which are both considered a cancer diagnosis.(3, 4) The number of breast cancer diagnoses has increased since the introduction of screening, due to both lead time and changes in underlying risk. In a mature cervical cancer screening program, the screen-detection of invasive cancer is rare due to the higher frequency of detection of precursor lesions, thus altering the natural history of those lesions that are progressive. Screening for cervical cancer mostly detects cervical intra-epithelial neoplasia (CIN). CIN is not regarded as a cancer diagnosis. The incidence rate of cervical cancer had been decreasing prior to the introduction of screening and has continued to decrease due to screening.(5, 6) To the degree that a colorectal cancer screening program focuses on the detection of adenomatous polyps and cancer, incidence also is expected to decline along with mortality after screening is introduced.(6)

A downside of early detection is the possibility of detecting abnormalities that would never have become clinically apparent in the absence of screening.(7) This may occur because abnormalities spontaneously regress, as is described for cervical cancer,(8-10) or that they remain indolent, as is described for breast cancer.(11, 12) Although there is little evidence to support the possibility of regression of breast cancer, it has been shown in vitro.(13) The detection of such an abnormality is called overdiagnosis, and most overdiagnoses lead to overtreatment. Overdiagnosis has been the topic of a fierce debate in breast cancer screening.(7) In cervical cancer screening, overdiagnosis is usually quantified as a lack of specificity for clinically significant disease.

The impact of overdiagnosis depends on its frequency and its consequences. In breast cancer screening, the overdiagnosis rate is relatively low,(7, 14, 15) in cervical cancer screening the overdiagnosis rate as such is usually not established. The consequence of overdiagnosis is unnecessary treatment which is inherently harmful. The consequences of overdiagnosis in breast cancer screening are more severe than those in overdiagnosed non-progressive CIN in cervical cancer screening. For an individual patient the name of the disease carries weight as well.

We aimed to exemplify the impact of overdiagnosis by comparing these two screening programs, which have been implemented for several decades in the Netherlands; for cervical cancer since 1985 for women aged 30-60 every five years, and for breast cancer since 1990 for women aged 50-74 every two years.(16, 17)

Estimates of the measure of overdiagnosis in breast cancer screening in literature vary from 4-54%.(7, 14, 18-20) The proper estimate of overdiagnosis has been the topic of many debates and the cause of many misunderstandings. We chose to use the definitions put forward by the UK independent review panel.(21)

This is the first simulation study aiming to compare different screening programs by addressing the potential amount and composition of overdiagnosed cases in the same overdiagnosis framework.

MATERIALS AND METHODS

The MISCAN model is a microsimulation model. This means the model simulates all individual life histories in a population. We have a model for breast cancer screening (MISCAN-Breast) and a model for cervical cancer screening (MISCAN-Cervix).(7, 22) In order to obtain a representative population, the models are fitted with a birth table and a life table. Each life history has its own probability of developing a (pre-) cancerous lesion. In MISCAN-Breast this probability is determined by fitting the model parameters: hazard, onset, and incidence, to data on incidence without screening from the Netherlands Cancer Registry. In MISCAN-Cervix the model is fitted to incidence data from Dutch Cancer Registry Database and data on detection from PALGA/PALEBA.(23)'(24) From each state the disease may progress to the next stage by a semi-Markov progression model (Figure 1). In MISCAN-Breast screening is implemented in the model using data on gradual roll-out, attendance rate and re-attendance rate in the Dutch screening program. Sensitivity, stage distribution, distribution of sojourn-time were estimated by fitting these parameters to data on incidence and stage distribution with screening (1991-2010) and without screening (1990). MISCAN-Breast assumes a 1.4% annual percentage change in underlying incidence.(25) Mortality reduction in the breast cancer model is based on the results of the Swedish trials.(26) The mortality reduction in the cervical cancer model is based on observational data, provided by the Dutch Cancer Registry and PALGA in the years 1998-2007.

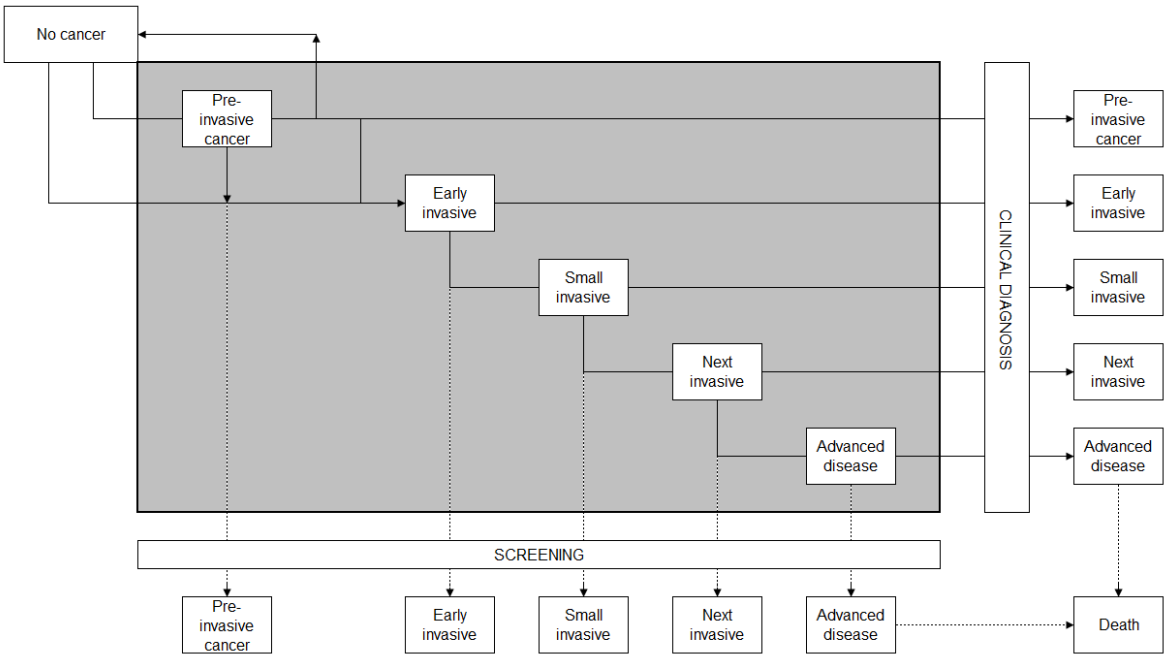


Figure 1. Progression in the MISCAN model. Every woman starts at the top left, where she has no cancer. From there she may progress through the different stages of cancer. If the cancer is detected by screening, the woman moves to the bottom of the graph (screen-detected). If the cancer is clinically detected she moves to the far right of the graph (clinically detected).

The impact of screening on an individual life history is illustrated by Figure 2, in which there are five different women, and each has a scenario without (A) and with (B) screening. The black areas are the negative effects of screening (life years with lower quality due to diagnosis and treatment), and dark grey areas are the positive effects of screening (healthy life years gained). Woman number 1 will benefit from screening. In situation 1A, there is no mass screening. She will have an onset of cancer; this cancer will grow and develop up to the point when she develops symptoms. The cancer will be clinically diagnosed and she will die from this cancer. In situation 1B, there is mass screening. The woman will have the same onset and the same preclinical disease phase, but now mass screening will detect her cancer before she develops symptoms. Therefore the disease is in a less advanced state and treatment is successful. She has gained life-years and will die of other causes than cancer. Woman number 2 does not benefit from screening. Like woman number 1, she has an onset of cancer, followed by a preclinical disease phase. This phase however, would extend beyond her lifespan. She will never be diagnosed with cancer in the situation without screening (2A). In the situation with screening (2B) the cancer will be detected by screening and she will be treated accordingly. She will still die at the same time, but now she has lost several quality-adjusted-life-years (QALY) due to the fact that she had a cancer diagnosed.

Woman number 3 develops a pre-invasive disease that will progress to a clinically detected cancer, but she will not die from this cancer (3A). She will also not gain any life-years by screening (3B). Woman number 4 has a type of cancer with an obvious pre-invasive precursor state (i.e. CIN in cervical cancer). In this case the preclinical phase is divided into two phases, one with preclinical pre-invasive disease and one with preclinical cancer. The preclinical-pre-invasive state will progress to preclinical cancer, which becomes clinically detected and leads to cancer-related death in the situation without screening (4A). When this woman is screened (4B) while her disease is in the pre-invasive phase and her condition is detected, she may be cured completely and thus cancer was prevented, she benefits from screening. Woman number 5 does not benefit from screening; she has a preclinical-pre-invasive disease that will not progress, or may even regress back to normal without screening (5A). Screening (5B) will give her a diagnosis of pre-invasive disease, but she will not gain any life-years.

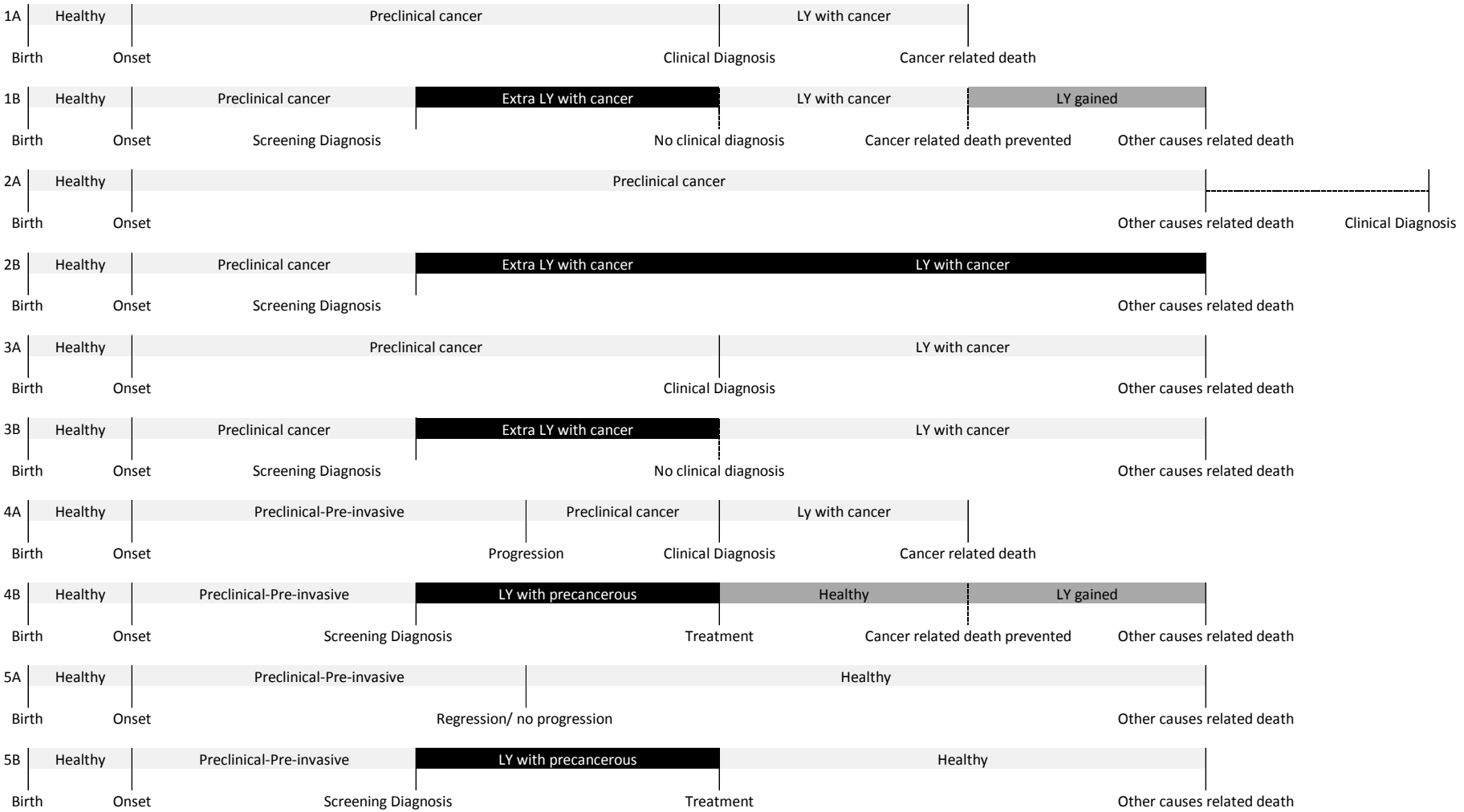


Figure 2. Life histories of women affected differently by screening. The numbers indicate different women, each of them having a life history without screening (A) and with screening (B). LY= life-years. Black areas represent negative effects of screening (overdiagnosis), dark grey areas represent the positive effects of screening (life years gained).

MISCAN-breast assumes a regression rate of 2%, and a progression rate of 11%, for DCIS.(27) MISCAN-Cervix has six different disease paths, five assume regression, and assumes progression from onset to invasive disease. Each woman has an age-dependent probability of ending up in one of the disease paths.

We performed a cohort run using our breast cancer model and cervical cancer model. The cohort consisted of 10,000,000 women, all born in 1982. The year 1982 was chosen so all women were 30 years and invited for cervical cancer screening in 2012, the most recent year with complete data. The number of simulated women alive in 2012 was also chosen as the denominator to convert raw data to rates. Between 2012 and 2032 (the year all women are invited to breast cancer screening for the first time) approximately 2% of the simulated women die of all-cause mortality (including cancer). Follow-up was completed for ages 30-100 years. Output measures were: number of diagnoses during entire follow-up in the situation without screening and in the situation with screening, and the number of diagnoses during the screening ages in the situation without screening and in the situation with screening. All results are presented per 100,000 women aged 30 in 2012 and stratified by pre-cancer (DCIS for breast cancer and CIN grades I, II and III for cervical cancer) and invasive cancer.

To estimate overdiagnosis we used the definitions set forward by the UK Independent review panel, which are: 1. “from the population perspective, the proportion of all cancers ever diagnosed in women invited to screening that are overdiagnosed”, and 2. “from the perspective of a woman invited to screening, the probability that a cancer diagnosed during the screening period represents overdiagnosis”.(21) To be able to address all diagnoses in the program, we extend the definitions above to include pre-invasive lesions, such as CIN I, II and III.

These definitions translate into the following calculations:

1. From the population perspective: Number of extra diagnoses with screening/Total number of diagnoses in a population with screening. For the purpose of comparison we used ages 30-100 years. No significant amount of cancers occurs before the age of 30.

2. From an individual perspective: Number of extra diagnoses with screening/Total number of diagnoses in women of screening age. For breast cancer screening this age range is 49-75 years. For cervical cancer screening this age range is 29-60 years, but we used 29-64 years because the diagnostic process in cervical cancer screening may take some time due to follow-up. This definition corresponds to the risk of having an overdiagnosed cancer in the lifetime of screening.

The number of extra diagnoses with screening is the difference between the total number of diagnoses in women aged 0-100 without screening and the total number of diagnoses in women aged 0-100 with screening. When we consider overdiagnosis, we included pre-invasive disease. If we had not included pre-invasive disease, overdiagnosis measures would not have applied.

RESULTS

All results are given per 100,000 women aged 30 in 2012. The model predicted 1,669 cervical neoplasia diagnoses (Table 1) and 13,210 breast cancer diagnoses per 100,000 women without screening (Table 2). Screening added 3,999 cervical neoplasia diagnoses and 207 breast cancer diagnoses. The extra cervical cancer diagnoses were 4,920 extra CIN lesions, which cannot be clinically detected, and 921 (-55.2%) fewer cervical cancer diagnoses. The extra breast cancer diagnoses were the result of 376 extra DCIS diagnoses (+61.7%), and a 169 less invasive cancers (-1.3%). We stratified the number by CIN grade, because impact and treatment options differ for each grade.

From a population perspective, the breast cancer overdiagnosis rate estimate was 1.5%. Cervical cancer overdiagnosis rate estimates from a population perspective varied from 70.6%, when including all CIN and invasive diagnoses to 50.0%, when including only CIN III and invasive disease. From the individual perspective, the breast cancer overdiagnosis rate estimate was 2.5%. Cervical cancer overdiagnosis rate estimates from this perspective varied from 74.8%, when including all CIN and invasive diagnoses, to 55.4%, when including only CIN III and cervical cancer (Table 3).

For women aged 30-100 we predicted 266 cervical cancer deaths with screening and 644 without screening , a mortality reduction of 59%. For women aged 30-100 we predicted 3,668 breast cancer deaths with screening and 4,637 without screening , a mortality reduction of 21%.

Table 1. Cervical cancer. Number of cases by stage and overdiagnosis rate per 100,000 women aged 30 years in 2012. CIN= cervical intra epithelial hyperplasia.

	No screening	With screening		
		Both clinically and screendetected	Screendetected	Clinically detected
Diagnoses during entire life (Ages 30-100 years)				
CIN I	0	1,138	1,138	0
CIN II	0	1,189	1,189	0
CIN III	0	2,593	2,593	0
Cervical cancer	1,669	748	117	632
Total	1,669	5,668	5,037	632
Diagnoses during screening (Ages 30-64 years)				
CIN I	0	1,138	1,138	0
CIN II	0	1,189	1,189	0
CIN III	0	2,593	2,593	0
Cervical cancer	1,138	424	117	307
Total	1,138	5,344	5,037	307
Cervical cancer deaths (Ages 30-100)	644	266		
Mortality reduction		59%		

Table 2. Breast cancer. Number of cases by stage and overdiagnosis rate per 100,000 women aged 30 years in 2012. DCIS= ductal carcinoma in situ.

	No screening	With screening		
		Both clinically and screendetected	Screendetected	Clinically detected
Diagnoses during entire life (Ages 30-100 years)				
DCIS	610	985	531	454
Breast cancer	12,600	12,432	3,523	8,908
Total	13,210	13,417	4,055	9,362
Diagnoses during screening (Ages 49-75 years)				
DCIS	364	746	531	215
Breast cancer	7,286	7,447	3,523	3,924
Total	7,650	8,194	4,055	4,139
Breast cancer deaths (Ages 30-100)	4,637	3,668		
Mortality reduction		21%		

Table 3. Overdiagnosis in Cervical cancer and Breast cancer. Number of cases by stage and overdiagnosis rate per 100,000 women aged 30 years in 2012. Excess diagnoses were calculated by subtracting all diagnoses in women aged 30-100 in the situation without screening from all diagnoses in women aged 30-100 in the situation with screening. Lifetime diagnoses are all diagnoses in women aged 30-100. Screening age diagnoses are all diagnoses in women aged 30-64 for cervical cancer, and in women aged 49-75 for breast cancer. CIN= cervical intra epithelial hyperplasia. DCIS= ductal carcinoma in situ.

Population perspective		Individual perspective	
Overdiagnosis rate =	Excess diagnoses / Lifetime diagnoses	Overdiagnosis rate=	Excess diagnoses / Screening age diagnoses
CIN I+II+III+cervical cancer	70.6%		74.8%
CIN II+III+cervical cancer	63.2%		68.0%
CIN III+cervical cancer	50.0%		55.4%
DCIS + breast cancer	1.5%		2.5%

DISCUSSION

Our comparison of the burden of breast cancer screening to that of cervical cancer screening shows that screening prevents cancer specific mortality, but when also including the detection of pre-invasive lesions in this equation, both types of screening also generate overdiagnosis.

The burden of overdiagnosis depends on its frequency and its consequences. Although the overdiagnosis frequency is high in cervical cancer screening relative to breast cancer screening, the impact is limited because treatment is minimally invasive. For CIN I most often no treatment is necessary, and for CIN II or III a loop excision or conisation may be done in an out-patient setting.(28) These procedures have relatively limited risks, and no apparent cosmetic impact. However, cold knife conisation and large loop excision may be associated with preterm delivery, low birth weight, caesarean section and preterm rupture of the membranes in future pregnancies.(29-31) For breast cancer screening, the frequency is low relative to cervical cancer screening, but the impact is higher due to more invasive treatment. The treatment of DCIS is lumpectomy or even mastectomy, in some cases followed by radiation therapy.(32, 33) The risks of these treatments include (rare) standard operation risks (haemorrhage or infection), and the risk of generalized anaesthesia. Additionally the cosmetic result of these procedures has significant impact.(33) The perception of the individual also needs to be taken into account. The information provided with each diagnosis, whether it is cancer or pre-invasive disease is crucial to the impact of this event.

The decision to count a diagnosis as overdiagnosis has to be related to its severity, treatment warranted, and on the impact of the information provided at diagnosis.

Our estimates for overdiagnosis of breast cancer were different from those previously published using the MISCAN model. This is a direct result of using cohort runs instead of simulating a realistic population. If we run our model with a population aged 0-100, we obtain an overdiagnosis rate directly comparable to that of De Gelder et al.(7) This rate is: from a population perspective, for all diagnoses 4.6%; and from an individual perspective, for all diagnoses 8.1%.(7, 14, 18-20) For cervical cancer no comparable numbers were published.

Our analysis for cervical cancer screening was performed on the current situation (i.e. primary conventional cytology testing with cytology triage) in the Netherlands. However, over the last years most laboratories have added a test to detect human papillomavirus (HPV) infections in the triage phase which slightly increases CIN I and CIN II detection.(34) In addition, most laboratories processing primary screening tests have switched from using conventional cytology to liquid-based cytology tests SurePath and ThinPrep. Rozemeijer et al. showed that CIN2+ detection rates increased by using SurePath, while they were

unaffected by using ThinPrep. this means that overdiagnosis rates are probably somewhat higher in the current Dutch situation than estimated in our study. Also, it is expected that from 2016 onwards, cervical cancer screening will be further modified in the Dutch program in 2016. Primary cytology will then be replaced by primary HPV screening with cytology triage. Furthermore, women will be invited for screening 5 times in their lifetime.(35) On the one hand there is a risk of increasing overdiagnosis by detecting disease at yet an earlier stage, on the other hand overdiagnosis may decrease due to less screening examinations in a life time.

Screening practices for breast and cervical cancer vary widely between countries. For example the United States do not have a national screening programme, though the recently updated American Cancer Society guideline for mammography screening recommends annual screening for women aged 45-54 and biennial screening after 55.(36) Cervical cancer screening in the US is carried out by many practitioners with shorter intervals than guidelines indicate, despite the recommendation made by the United States Preventive Services Task Force (USPSTF).(37, 38) Although the affordable Care Act now ties coverage to the USPSTF recommendations, more doctors follow the American Cancer Society (ACS) guidelines in breast cancer screening than the USPSTF recommendations. The NHS Breast cancer screening program in the UK invites women aged 50-70 every three years and is currently extending to include women aged 47-73. (39).In cervical cancer screening in the United States the guidelines are similar to those in the Netherlands.(40, 41) In another example the cervical cancer screening program of Finland is comparable to that of the Netherlands (but with considerably more opportunistic screening), while in the UK, Sweden and Denmark women are screened 12 and 13 times a lifetime starting at the ages of 25, 23 and 23, respectively.(42-44) Therefore our estimates may be different in other countries. Overdiagnosis estimates are expected to increase for both screening programs with increasing number of screening examinations and with a younger age at first screening. Every early diagnosis can lead to overdiagnosis, because other cause mortality may occur before the benefits of early treatment are realized. This is most likely in older women, but it may also occur in younger women, especially with indolent disease In fact more non-progressive CIN is found in younger women than in older women. (45)

Looking towards the future, if we were to analyse the data for colorectal cancer screening we would expect results in between those of breast cancer and cervical cancer screening, depending on the screening test being used. Faecal occult blood tests, especially the older guaiac tests but also the newer e.g. immunochemical tests, have a lower sensitivity for early, pre-invasive disease than endoscopy. The most sensitive test will find more pre-invasive disease, which will need less invasive treatment but also more often would not have developed into clinical disease, so the frequency of overdiagnosis would be high but the per case consequences would be low.

As more types of cancer will become eligible for screening, we hope that in the future balanced reports will elucidate the impact of any cancer screening on the advanced cancer rate and disease specific mortality while also publishing the properly estimated extent of overdiagnosis.

LIMITATIONS OF THE STUDY

In order to compare the two programs, which offer screening at different ages, we performed a cohort run. Although this results in a lifetime estimate of harms and benefits, it remains hypothetical as the homogeneity of a cohort never resembles a real population. Mathematical modelling requires assumptions made in the model on natural history of cancer. The mean duration of sojourn time and the probability of progression are interchangeable in the model, the assumptions used have influenced the overdiagnosis estimate. (46)

We have extended the definition of overdiagnosis somewhat by including precursor lesions that commonly are not judged to be cancers. This is not the case so much with DCIS of the breast, but precursor lesions of the cervix have not commonly been included in discussions of overdiagnosis, nor have adenomas. In the case of each, the fraction of overdiagnosis will be harder to estimate because you would need to estimate the fraction of treated lesions that were progressive, something that is quite uncertain.

CONCLUSION

We have compared the burden of screening for two of the population screening programs for cancer currently in use in the Netherlands. For breast cancer, overdiagnosis estimates are relatively low, but the consequences for overdiagnosed women are significant. For the program overall these consequences however are quite small. For cervical cancer, overdiagnosis estimates of pre-invasive disease are high, but the consequences are relatively small due to less invasive treatment. Informing women about the potential harms of screening should include the consequences of finding the different lesions, invasive or pre-invasive.

REFERENCES

1. Wilson JMGJ, G. Principles and Practice of Screening For Disease. Geneva: WHO; 1968.
2. Boer R, Warmerdam P, de Koning H, van Oortmarssen G. Extra incidence caused by mammographic screening. Lancet. 1994;343(8903):979.
3. Burstein HJ, Polyak K, Wong JS, Lester SC, Kaelin CM. Ductal carcinoma in situ of the breast. N Engl J Med. 2004;350(14):1430-41.
4. Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ. JAMA Oncol. 2015;1(7):888-96.
5. Bulk S, Visser O, Rozendaal L, Verheijen RH, Meijer CJ. Cervical cancer in the Netherlands 1989-1998: Decrease of squamous cell carcinoma in older women, increase of adenocarcinoma in younger women. Int J Cancer. 2005;113(6):1005-9.
6. Esserman LJ, Thompson IM, Jr., Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. JAMA. 2013;310(8):797-8.
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(1):111-21.
8. McCredie MR, Sharples KJ, Paul C, Baranyai J, Medley G, Jones RW, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. Lancet Oncol. 2008;9(5):425-34.
9. Bansal N, Wright JD, Cohen CJ, Herzog TJ. Natural history of established low grade cervical intraepithelial (CIN 1) lesions. Anticancer Res. 2008;28(3B):1763-6.
10. Moscicki AB, Ma Y, Wibbelsman C, Darragh TM, Powers A, Farhat S, et al. Rate of and risks for regression of cervical intraepithelial neoplasia 2 in adolescents and young women. Obstet Gynecol. 2010;116(6):1373-80.
11. Page DL, Dupont WD, Rogers LW, Jensen RA, Schuyler PA. Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer. 1995;76(7):1197-200.
12. Albrektsen G, Heuch I, Thoresen SO. Histological type and grade of breast cancer tumors by parity, age at birth, and time since birth: a register-based study in Norway. BMC Cancer. 2010;10:226.
13. Warenius H, Kyritsi L, Grierson I, Howarth A, Seabra L, Jones M, et al. Spontaneous regression of human cancer cells in vitro: potential role of disruption of Cdk1/Cdk4 co-expression. Anticancer Res. 2009;29(6):1933-41.
14. Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. Breast Cancer Res. 2005;7(6):258-65.
15. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56.
16. van Ballegooijen M, Hermens R. Cervical cancer screening in the Netherlands. Eur J Cancer. 2000;36(17):2244-6.
17. Otten JD, van Dijck JA, Peer PG, Straatman H, Verbeek AL, Mravunac M, et al. Long term breast cancer screening in Nijmegen, The Netherlands: the nine rounds from 1975-92. J Epidemiol Community Health. 1996;50(3):353-8.

18. Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. Cancer Causes Control. 2010;21(2):275-82.
19. Welch HG, Black WC. Overdiagnosis in cancer. J Natl Cancer Inst. 2010;102(9):605-13.
20. 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(7):491-9.
21. 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(11):2205-40.
22. De Kok IM, van Rosmalen J, Van Ballegooijen M. Description of MISCAN-cervix, Web Appendix accompanying 'A comparison of primary HPV to cytology cervical cancer screening in different European settings: A costeffectiveness analysis based on a Dutch microsimulation model'. RePub EUR; [09/03/2013]. Available from: http://repub.eur.nl/res/pub/31582/deki874693.ww1.pdf.
23. Cijfers over Kanker: The Netherlands Cancer Registry; 2014. Available from: http://www.cijfersoverkanker.nl/.
24. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol. 2007;29(1):19-24.
25. Gelder Rd. Predicting the Benefits and Harms of Breast Cancer Screening: Current debates and future directions. Rotterdam: Erasmus University; 2012.
26. Tabar L, Yen MF, Vitak B, Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet. 2003;361(9367):1405-10.
27. de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: Weighing reduced mortality against increased overdiagnosis. Prev Med. 2011.
28. Gram IT, Bremnes Y, Ursin G, Maskarinec G, Bjurstam N, Lund E. Percentage density, Wolfe's and Tabar's mammographic patterns: agreement and association with risk factors for breast cancer. Breast Cancer Res. 2005;7(5):R854-61.
29. Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. Lancet. 2006;367(9509):489-98.
30. Verbeek AL, Broeders MJ, Otto SJ, Frachebou J, Otten JD, Holland R, et al. [Effects of the population screening into breast cancer]

Effecten van het bevolkingsonderzoek naar borstkanker. Ned Tijdschr Geneeskd. 2013;157(10):A5218.
31. Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. Radiology. 2012;265(3):707-14.
32. NABON. Breast cancer, Dutch guideline, version 2.0: oncoline; 2012 [09/03/2013]. Available

from: http://www.oncoline.nl/uploaded/docs/mammacarcinoom/Dutch%20Breast%20Cancer%20Guideline%202012.pdf.
33. McLaughlin SA. Surgical management of the breast: breast conservation therapy and mastectomy. Surg Clin North Am. 2013;93(2):411-28.
34. Siebers AG, Arbyn M, Melchers WJ, van Kemenade FJ, Vedder JE, van der Linden H, et al. Effectiveness of two strategies to follow-up ASC-US and LSIL screening results in The Netherlands using repeat cytology with or without additional hrHPV testing: a retrospective cohort study. Cancer Causes Control. 2014;25(9):1141-9.
35. Population screening for cervical cancer. The Hague2011.
36. Oeffinger KC, Fontham ET, Etzioni R, Herzig A, Michaelson JS, Shih YC, et al. Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. JAMA. 2015;314(15):1599-614.
37. Mitka M. Physicians, patients not following advice from USPSTF on mammography screening. JAMA. 2013;309(20):2084.
38. Berkowitz Z, Saraiya M, Sawaya GF. Cervical cancer screening intervals, 2006 to 2009: moving beyond annual testing. JAMA Intern Med. 2013;173(10):922-4.
39. ClinicalTrials.gov. ClinicalTrials.gov, A service of the U.S. National Institutes of Health 2011 [cited 2014 28-01-2014]. Available from: http://clinicaltrials.gov/ct2/show/NCT01081288.
40. Yabroff KR, Saraiya M, Meissner HI, Haggstrom DA, Wideroff L, Yuan G, et al. Specialty differences in primary care physician reports of papanicolaou test screening practices: a national survey, 2006 to 2007. Ann Intern Med. 2009;151(9):602-11.
41. Lee J, Gordon PB, Whitman GJ. "Do unto others as you would have them do unto you": breast imagers' perspectives regarding screening mammography for others and for themselves--do they practice what they preach? AJR Am J Roentgenol. 2015;204(6):1336-44.
42. Dillner J. Cervical cancer screening in Sweden. Eur J Cancer. 2000;36(17):2255-9.
43. Denmark NboH. Cervical cancer screening. http://sundhedsstyrelsen.dk/publ/Publ2007/PLAN/Kraeft/Anbef_screen_livmoderhals_en_samftn.pdf: Sundhedsstyrelsen; 2007. 7 p.
44. NHS. NHS cervical screening programme: Public Health England; 2013 [cited 2014 28-01-2014]. Available from: http://www.cancerscreening.nhs.uk/cervical/about-cervical-screening.html.
45. van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer. 1991;64(3):559-65.
46. de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med. 2011;53(3):134-40.

CHAPTER 4:

THE DISTRIBUTION OF DCIS GRADE IN 4,232 WOMEN AND ITS IMPACT ON OVERDIAGNOSIS IN BREAST CANCER SCREENING

Paula van Luijt

Eveline Heijnsdijk

Jacques Fracheboud

Lucy Overbeek

Mireille Broeders

Jelle. Wesseling

Gerard den Heeten

Harry de Koning.

Breast Cancer Res. 2016;18(1):47.

ABSTRACT

INTRODUCTION: The incidence of ductal carcinoma in situ (DCIS) has rapidly increased over time. The malignant potential of DCIS is dependent on its differentiation grade.

MATERIALS AND METHODS: Our aim is to determine the distribution of different grades of DCIS among women screened in the mass-screening program, and women not screened in the mass screening program and to estimate the amount of overdiagnosis by grade of DCIS. We retrospectively included a population-based sample of 4,232 women with a diagnosis of DCIS in the years 2007-2009 from the Nationwide network and registry of histopathology and cytopathology in the Netherlands. Excluded were women with concurrent invasive breast cancer, lobular carcinoma in situ and no DCIS, women recently treated for invasive breast cancer, no grade mentioned in the record, inconclusive record on invasion, and prevalent DCIS. The screening status was obtained via the screening organizations. The distribution of grades was incorporated in the well-established and validated microsimulation model MISCAN.

RESULTS: Overall, 17.7% of DCIS were low-grade, 31.4% intermediate-grade, and 50.9% high-grade. This distribution did not differ by screening status, but did vary by age. Older women were more likely to have low-grade DCIS than younger women. Overdiagnosis as a proportion of all cancers in women of the screening age was 61% for low-grade, 57% for intermediate-grade, 45% for high-grade DCIS. For women age 50-60 years with a high-grade DCIS this overdiagnosis rate was 21-29%, compared to 50-66% in women age 60-75 years with high-grade DCIS.

CONCLUSION: Amongst the rapidly increasing numbers of DCIS diagnosed each year is a significant number of overdiagnosed cases. Tailoring treatment to the probability of progression is the next step to preventing overtreatment. The basis of this tailoring could be DCIS grade and age.

INTRODUCTION

Ductal carcinoma in situ (DCIS) is a “neoplastic proliferation of cells within the ductal-lobular structures of the breast that has not penetrated the myoepithelial-basement membrane interface”.(1) Before the introduction of mammography screening, DCIS was rarely diagnosed. In 1989, 366 women in the Netherlands were diagnosed with DCIS. In 2003, more than 10 years after the introduction of mass-screening, 1,171 women had a DCIS diagnosed. With the introduction of digital screening this figure rose to 2,046 women in 2011, and most recently to 2,406 in 2014.(2)

The extent to which DCIS represents overdiagnosis has been extensively debated in relation to organised screening programmes.(3-6) Overdiagnosis is defined as a lesion diagnosed by screening in an asymptomatic woman that would not have been detected during the woman’s lifetime in the absence of screening.(4) To predict the probability of a DCIS to progress to invasive carcinoma, six different grading systems were proposed, based on morphology or molecular profile.(7) All of these classify DCIS into three categories of malignant potential: Low (I), intermediate (II), or high (III). The grade of DCIS is correlated with the risk of progression, as well as with the grade of concurrent invasive carcinoma.(8-13) The transition from low-grade DCIS to high-grade DCIS or to high-grade invasive carcinoma is deemed unlikely.[8-10, 12]

The grade distribution of DCIS has been studied in mostly small series,(6, 14-18) or only included screen-detected cases (Table 1). (19) More insight in this distribution based on larger numbers in both screened and non-screened populations is of paramount importance and may improve our estimates of overdiagnosis.

The aim of this study was to establish the distribution of different grades of DCIS in different subgroups based on mass-screening status and age group, and to estimate the overdiagnosis rate for each grade and age group specifically.

Table 1. Grade distribution by detection mode and sample size compared to previous studies on DCIS grade distribution.

	Screendetected				Symptomatic				Not specified			
	DCIS grade 1 %	DCIS grade 2 %	DCIS grade 3 %	N	DCIS grade 1 %	DCIS grade 2 %	DCIS grade 3 %	N	DCIS grade 1 %	DCIS grade 2 %	DCIS grade 3 %	N
This study	16	32	52	1430	19	28	54	263				
Evans 2001	13	18	69	222	16	23	61	151	-	-	-	-
Kessar 2002	23	23	54	98	19	19	62	52	-	-	-	-
Meijnen 2005	22	31	47	87	26	24	49	293	-	-	-	-
De Roos 2007	7	44	53	54	30	45	25	20	-	-	-	-
Sorum 2010	-	-	-	-	-	-	-	-	23	23	53	2403
Bluekens 2012	15	32	54	853	-	-	-	-	-	-	-	-
Weigel 2015	18	39	42	898	-	-	-	-	-	-	-	-

MATERIALS AND METHODS

PATIENT SELECTION

We obtained 17,744 excerpts from 12,301 women with DCIS from the years 2007, 2008 and 2009 from the 'Nationwide network and registry of histopathology and cytopathology in the Netherlands' (PALGA). PALGA is a national database containing the excerpts and coded diagnoses of all pathological and cytological examinations performed in the Netherlands.(20) The mass-screening status of these women was established by linking the database to the databases of the screening organizations by an independent third party, with permission of the screening organizations. Our database contained anonymized records of mass-screening status (positive, negative, year of last mass-screening and number of mass-screening examinations), age, year of diagnosis, and a short summary of the conclusion of the original pathology report.

From the 12,301 women, we excluded those who also had a concurrent invasive breast cancer (ipsilateral or contralateral, N=7,089), those who had a lobular carcinoma in situ and no DCIS (N=6), those who turned out after excision biopsy or ablation not to have any malignancy (N=131), those who had recently been treated for invasive breast cancer (N=247), those who had no grade mentioned in the excerpt (N=17), those who had an inconclusive excerpt on invasion or otherwise (N=242), and women who had a prevalent DCIS, rather than a new diagnosis in the study period (N=354). We excluded contralateral disease because our model does not include bilateral disease.

DCIS DETECTED BY MASS-SCREENING

DCIS were assumed to be 'detected by mass-screening' when a woman had had a positive screening examination in between 2007 and 2009. Women who had participated in the screening program, and did not have any positive screens, but who did have a DCIS diagnosis in 2007, 2008, or 2009 were assumed to have an interval-DCIS. The number of interval-DCIS increased across the study period due to the cumulative effect of interval-DCIS diagnosed in women screened in the previous year (2007) or in the two previous years (2007 and 2008). Interval-DCIS were rare in 2007 because of the low frequency of interval carcinoma's within the same calendar year in which the screening examination took place. In 2008 interval-DCIS were diagnosed in women screened in 2007 or 2008, and in 2009 interval-DCIS were diagnosed in women screened in 2007, 2008 or 2009.

Women who were not known to the screening organizations may have been under clinical surveillance because of high familial risk, frequent (benign) breast anomalies, or because of personal preference. Diagnoses in this group may be the result of screening, but are not the result of the mass-screening program. Therefore we cannot conclude that DCIS not detected by mass-screening, were not detected by screening. To compare the distribution of DCIS

detected by mass-screening to DCIS not detected by mass-screening, we, therefore, chose to compare the DCIS detected by mass-screening to the interval-DCIS.

GRADING OF DCIS

In line with the Dutch guidelines, the classification by Holland et al. is almost exclusively used.(21) At the start of the mass-screening program in the early 1990's, pathologists were instructed on how to uniformly classify each DCIS.

DCIS grade was determined using the information in the short summary of the pathology report by description, i.e. high, moderate, or low differentiation; low, intermediate, or high malignancy potential; or grade I, II, or III. If the summary contained more than one grade this case was graded according to the highest grade mentioned. If there was a discrepancy between grades in different specimens of the same patient, the grade was based on the most representative specimen, i.e. resection is more representative than biopsy, but biopsy is more representative than cytology.

STATISTICAL ANALYSIS

Proportions of DCIS grades were calculated by year, age group, and screening status. We compared these proportions between screening groups using the Pearson chi-square test. Multivariate analyses on age groups were performed with a logistic regression model. The statistically significant parameters were identified by the introduction of variables in a stepwise manner. All calculations were performed using SPSS 20.

MODELLING APPROACH

The MISCAN model is a microsimulation model that simulates the individual life histories of women.(22) The probability of each woman to have an onset of breast cancer is determined by calibrating the model to the incidence rate in 1989 (the year before screening was introduced), adjusted with an annual percentage change of 1.4% to account for the rising background breast cancer incidence.(23) The natural history of breast cancer is modelled as a Markov-like progression through the successive pre-clinical stages of the disease. Details of the model have been described previously.(4) For this analysis we added the three DCIS grades to the model, using the age-dependent grade distribution found in this study (Figure 1).

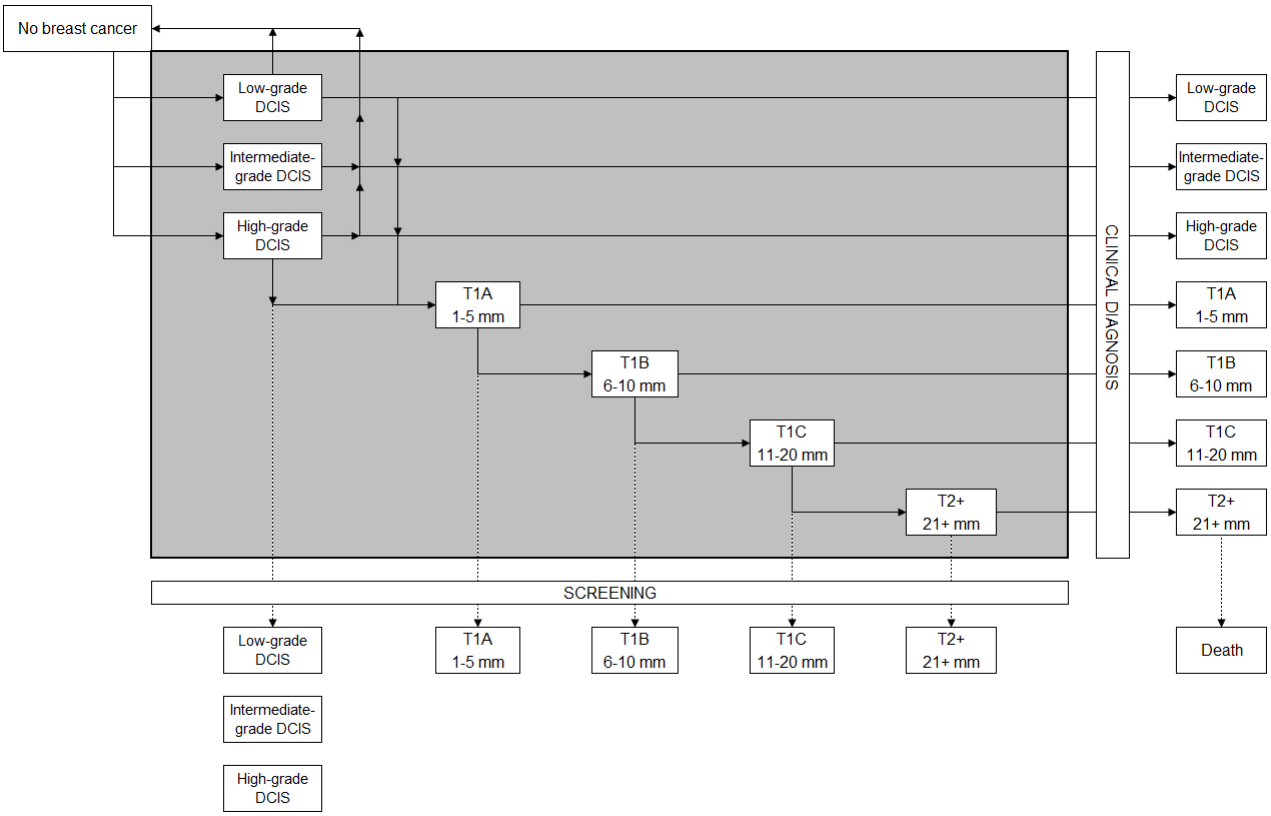


Figure 1. Schematic drawing of the extended MISCAN model. Transition possibilities are indicated with arrows. All diseases within the grey area are preclinical disease, after diagnosis they are either clinically detected or detected by mass-screening. DCIS= Ductal Carcinoma In Situ, T1a= tumour with a diameter up to 5 mm, T1b= tumour with a diameter from 5 mm up to 10 mm, T1c= tumour with a diameter from 10 mm up to 20 mm, T2+= any tumour with a diameter larger than 20 mm. There is no transition between low-grade DCIS, intermediate-grade DCIS and high-grade DCIS.

Following onset, breast cancer in a pre-clinical stage can progress to the next pre-clinical stage (dependent on the duration of the previous state), or become clinically detected. In addition, the DCIS stages may also regress to normal.(24, 25) Screening is superimposed on this life history.

The transition probabilities, duration of tumour stages, and test sensitivities were calibrated using data from the Dutch population and Dutch breast cancer screening from 1975 to 2010 on breast cancer incidence by stage, age, and detection mode. The Dutch nationwide breast-cancer screening program has invited all women aged 50-69 since 1990 and women aged 50-75 since 1998 biennially for a mammographic screening examination, free of charge. The attendance rate is approximately 80%.(26)

We chose to look at model-outcomes for the years 2000-2009 because there was a steady state situation in these years, more than 10 years after the start of the screening program. We evaluated the following output: incidence rate by detection mode (screen detected or

clinically detected), age, and year of diagnosis. The model compares women in the situation with screening, to the same women in the situation without screening; if a woman has a screen-detected cancer, but would not have had a diagnosis in the situation without screening, this case is regarded as overdiagnosed (Figure 2).

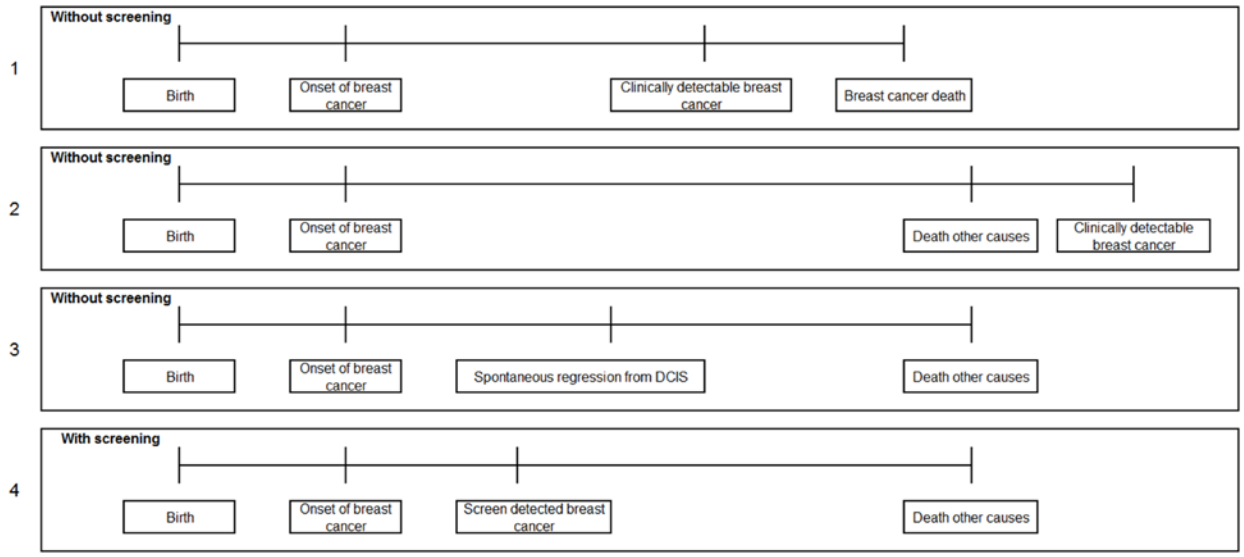


Figure 2. Screening affecting three women differently. The first box is the life history of a woman who has an onset of breast cancer, is diagnosed clinically, and dies of breast cancer. The second box is the life history of a woman who also has an onset of breast cancer, but who dies of other causes before this would be detected. The third box is the life history of a woman who has an onset of breast cancer, but also a spontaneous regression, this woman would not have been diagnosed without screening. The fourth box indicates the situation for these three women had screening been introduced. The woman in the first box no longer dies from breast cancer; the other two women do not benefit from screening, they have been overdiagnosed.

The estimates and definitions of overdiagnosis vary widely among international publications. [(4)] To minimize confusion we used the definitions of overdiagnosis which were deemed most useful by an independent review panel in the UK; from a population perspective: The proportion of all cancers ever diagnosed in women of the screening age and over (50-100 years) that are overdiagnosed; and from an individual perspective: The proportion of all cancers ever diagnosed in women of the screening age (50-75 years) that are overdiagnosed.(27)

ASSUMPTIONS ON NATURAL BEHAVIOUR OF DCIS

In the original model a 2% regression rate, a 11% progression rate, and a 5% clinical detection rate was assumed for all DCIS, resulting in a proper fit of incidence.(28) Little is known about the natural history of DCIS without treatment. Small studies were published, indicating a progression rate of 1 in 2 to 1 in 3 for low-grade DCIS, 1 in 3 for intermediate-grade DCIS and 2 in 3 in high-grade DCIS.(29, 30) Progression rate may differ from the rate assumed in the original model. In the new model we assumed that intermediate-grade DCIS has the same transition probabilities as all DCIS had in the original model. We lowered the regression rate to 1% for high-grade DCIS, and increased the regression rate to 4% for low-grade DCIS, based on the findings of Sanders et al.(30) The probability for a DCIS to be clinically detected was assumed independent of grade. The probability of progression: 16% for low-grade DCIS, 31% for intermediate-grade DCIS, and 53% for high-grade DCIS, was estimated by correcting the probabilities of low-grade DCIS and high-grade DCIS by the progression found in literature.(29, 30) Adjusting the progression rate and therefore the duration of the state, influences all successive states, because the progression of each successive state is dependent on the duration of the previous state. High-grade invasive breast cancer follows high-grade DCIS and low-grade invasive breast cancer follows low-grade DCIS. We calibrated DCIS incidence rate to observed data for the period 1990-2010. Parameters are summarized in table 2.

Table 2. Model parameters, sources and estimates.

Parameter		Estimate	Source
Incidence of breast cancer prior to screening	51.61 / 100,000 women		Dutch cancer registry
Current incidence of breast cancer	85.33 / 100,000 women		Dutch cancer registry
Grade distribution of DCIS	low grade DCIS	16.4% - 18.8%	PALGA (this study)
	intermediate grade DCIS	27.2% - 31.6%	
	high grade DCIS	52.0% - 54.0 %	
Progression rate	low grade DCIS	16%	Sanders (2005)
	intermediate grade DCIS	31%	
	high grade DCIS	53%	
Regression rate	low grade DCIS	4%	Collins (2005), Sanders (2005)
	intermediate grade DCIS	2%	
	high grade DCIS	1%	

RESULTS

PATIENTS/DISTRIBUTION OF DCIS GRADE

Patient characteristics are summarised in table 3. There was no significant difference in the distribution of grades between the DCIS detected by mass-screening and the DCIS not detected by mass-screening (from the interval group); 16.4-18.8% were low-grade, 27.2-31.6% were intermediate-grade, and 52.0-54.0% were high-grade (Table 4).

Univariate analysis of the group, not detected by mass-screening, showed that DCIS grade has an inverse linear association with 5-year age group (p-value=0.015), and with age as a linear variable (p-value=0.018). Year of diagnosis did not contribute in this group (Table 5). Overall the year of diagnosis was a significant independent variable (p-value=0.02).

ESTIMATING OVERDIAGNOSIS

The distribution of DCIS grade was included in the model and the new model was calibrated estimating dwell times and probabilities of transition on incidence data from the Cancer Registry and grade distribution from our study (Figure 3).

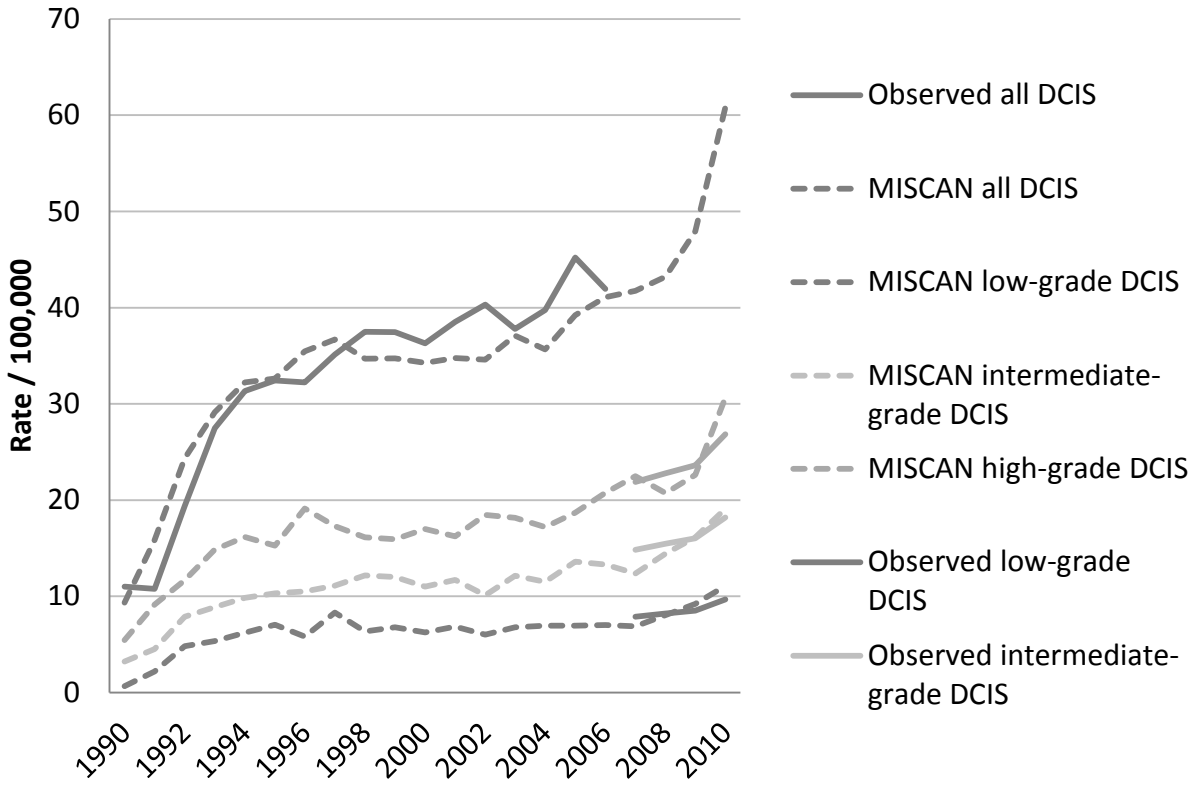


Figure 3. Low-grade DCIS, intermediate-grade DCIS and high-grade DCIS per 100,000 women aged 50-60. DCIS: Ductal Carcinoma In Situ. MISCAN: Predicted rates by the model. Observed: The number of DCIS as calculated when applying DCIS-grade distribution to the data on total DCIS incidence from the Dutch Cancer Registry.

Overdiagnosis estimates from the model were, from the population perspective; 60% of low-grade DCIS, 56% of intermediate-grade DCIS, 45% of high-grade DCIS. Overall (invasive disease and DCIS) overdiagnosis rate from the population perspective was 8%. Overdiagnosis estimates from the individual perspective were; 61% of low-grade DCIS, 57% of intermediate-grade DCIS, 45% of high-grade DCIS. When stratified by age group, the younger women had a much lower overdiagnosis rate when being diagnosed with a high-grade DCIS, varying from 21% in age group 50-55 to 29% in age group 55-60, up to 66% in age group 70-75 (Table 6).

Table 3. Descriptive statistics of the DCIS cases reviewed. 'Known at mass-screening' are all women who were listed in the database of the screening organisations, with a positive or a negative screen, 'Not known at mass-screening' are all women who were not mentioned in the screening organisation's database. N.a.: Not applicable.

	Known at mass-screening		Not known at mass-screening		P-value
	N	%	N	%	
Patients	4.075		8.226		
Exclusions	2.382	58%	5.687	69%	
Inclusions	1.693	42%	2.539	31%	
Year diagnosis					
2007	429	25%	865	34%	<0·001
2008	583	34%	806	32%	
2009	681	40%	868	34%	
Age group					
<49	0	0%	651	26%	<0·001
49-75	1.690	100%	1.686	66%	
>75	3	0%	202	8%	
Screen result					
Positive screen	1.430		n.a.		
No positive screen	263		n.a.		
	Mean		Mean		
Age	60·8		56·3		<0·001

Table 4. Distribution of different DCIS grades by screening status and age group. The p-values indicate the significance of the difference of these distributions between screening status. DCIS: Ductal Carcinoma In Situ. Low-grade DCIS: DCIS with a low malignant potential. Intermediate-grade DCIS: DCIS with an intermediate malignant potential. High-grade DCIS: DCIS with a high malignant potential.

		Detected at mass-screening		Screen negative		P-value
		N	%	N	%	
Age group						
<49		0		0		
	Low-grade DCIS	0	n.a.	0	n.a.	
	Intermediate-grade DCIS	0	n.a.	0	n.a.	
	High-grade DCIS	0	n.a.	0	n.a.	
49-75		1,429		261		
	Low-grade DCIS	234	16.4%	49	18.8%	
	Intermediate-grade DCIS	452	31.6%	71	27.2%	0.579
	High-grade DCIS	743	52.0%	141	54.0%	
>75		1		2		
	Low-grade DCIS	0	0.0%	0	0.0%	
	Intermediate-grade DCIS	0	0.0%	2	100.0%	0.297
	High-grade DCIS	1	100.0%	0	0.0%	

Table 5. Distribution of different DCIS grades by year and screening status. The p-values indicate the significance of the difference of these distributions between screening status. DCIS: Ductal Carcinoma In Situ. Low-grade DCIS: DCIS with a low malignant potential. Intermediate-grade DCIS: DCIS with an intermediate malignant potential. High-grade DCIS: DCIS with a high malignant potential.

		Detected at mass-screening		Screen negative		P-value
		N	%	N	%	
Year						
2007		410		19		
	Low-grade DCIS	59	14.4%	3	15.8%	
	Intermediate-grade DCIS	109	26.6%	2	10.5%	0.083
	High-grade DCIS	242	59.0%	14	73.7%	
2008		525		58		
	Low-grade DCIS	91	17.3%	11	19.0%	
	Intermediate-grade DCIS	167	31.8%	15	25.9%	0.827
	High-grade DCIS	267	50.9%	32	55.2%	
2009		495		186		
	Low-grade DCIS	84	17.0%	35	18.8%	
	Intermediate-grade DCIS	176	35.6%	56	30.1%	0.651
	High-grade DCIS	235	47.5%	95	51.1%	

Table 6. Overdiagnosis estimates by two different definitions. Population perspective: The proportion of all cancers ever diagnosed in women of the screening age and over (50-100 years) that are overdiagnosed. Individual perspective: The proportion of all cancers ever diagnosed in women of the screening age (50-75 years) that are overdiagnosed. DCIS: Ductal Carcinoma In Situ. Low-grade DCIS: DCIS with a low malignant potential. Intermediate-grade DCIS: DCIS with an intermediate malignant potential. High-grade DCIS: DCIS with a high malignant potential.

	Low-grade DCIS	Intermediate-grade DCIS	High-grade DCIS
Population perspective	60%	56%	45%
Individual perspective	61%	57%	45%
Individual perspective by age group			
50-55	58%	46%	21%
55-60	62%	55%	29%
60-65	66%	64%	50%
65-70	49%	52%	61%
70-75	54%	58%	66%

DISCUSSION

This is the largest study on the distribution of DCIS grade and the first modelling study to estimate overdiagnosis rate by DCIS grade. The distribution of grades in DCIS is dependent on age, but not on mass-screening status. This is in accordance with earlier studies on grade distribution. The overall distribution is also consistent with these studies (Table 4).(6, 14-16, 18, 19, 31)

The incidence rate of DCIS has increased rapidly over recent years. DCIS is unequivocally associated with mammography screening. Approximately one third of the cases in the database were detected by mass-screening, which corresponds to the overall distribution of breast cancers detected by mass-screening (both in situ and invasive) of all breast cancers in the Dutch population, and to the findings of Shin et al.(32) However, in our study, when linking Dutch pathology reports to the records of the screening organisations, most DCIS were not known at mass-screening organisations. This can partly be explained by the fact that one of the nine organisations, that were responsible for screening at the time, did not deliver data to be linked to the PALGA database. This organisation represents approximately 15% of all screened women annually. Secondly, we do not know how the diagnoses not detected by mass-screening were established. Given the age distribution and the fact that DCIS is generally not palpable, we assume that the majority of these cases are diagnosed through screening outside the mass-screening program.

As expected, and in line with previous studies, we found more low-grade DCIS in older women.(33) In general, more aggressive cancers are diagnosed earlier in life. Those that remain for detection at an older age are more likely to be less aggressive.(34)

In the Netherlands, a transition to screening with digital mammography was made between 2005 and 2010. In 2010, the detection rate of DCIS in mass-screening increased substantially, probably as a result of the introduction of digital mammography screening. Currently, it is not yet clear whether this is a prevalence effect or a lasting effect. We studied the years 2007, 2008 and 2009; thus, an increasing proportion of the DCIS we considered have been found with digital screening. We have no knowledge which DCIS were detected by digital mammography or film screen mammography. Also the DCIS detected outside the mass-screening program are equally likely to have been detected with digital mammography. We did not find a difference in grade distribution in screen detected DCIS over this period; therefore it seems unlikely that digital screening will have significantly altered the grade distribution, which is also in accordance with the findings of Bluekens et al.(19)

We have found that grade distribution for DCIS in the years 2007, 2008 and 2009, was inversely related to age, but we have no information on historical development of this distribution. For our study, we assumed the distribution to be stable over time.

Considerable controversy exists on whether DCIS is the ideal stage of the disease for early detection, or whether the detection of DCIS represents overdiagnosis, and, consequently, overtreatment. However, agreement exists that it is essential to determine which individual diagnosis is overdiagnosis and which is not. Central to this discussion is the natural behaviour of DCIS. Now that we have specified grade of DCIS in the microsimulation model, we can estimate overdiagnosis more accurately. Only 16.4% of detected by mass-screening DCIS are low-grade, 60% respectively 61% of which are overdiagnosed, depending on the definition of overdiagnosis. We found that 50.9% of all detected by mass-screening DCIS are high-grade, and therefore have a high risk of progression. In these cases we are bound to find aggressive cancer earlier and to prevent fast-growing invasive cancer, but even so, 45% of these cases are overdiagnosed, independent on the definition of overdiagnosis. For younger women (age 50-60) with a high-grade DCIS however, overdiagnosis estimates vary between 21% and 29% from an individual perspective, therefore for these women screening is most protective.

We found an increasing amount of overdiagnosis in older women with high-grade DCIS, this is the result of a longer dwell time in the model in high-grade DCIS in women over 60. This dwell time was calibrated by the model. A disease with a longer dwell time is more likely to be detected by screening. The longer dwell time of high grade DCIS in older women correlates to the findings of Weigel et al., who found a higher detection rate of high grade DCIS in older women.(33)

Our overdiagnosis estimates make a general decision on treatment from a population based approach a very difficult one for women with DCIS. We estimate that 60% of these women would be overtreated if they undergo treatment for this disease, of which they would never have been aware in the absence of screening. On the other hand, they are diagnosed with an entity that carries a specific risk for progression to an invasive and potentially lethal disease and will therefore lean towards treatment, rather than active surveillance. If this entity would be named differently this might be perceived differently.(35) DCIS can also be regarded as a risk factor like lobular carcinoma in situ. One can question whether the increased risk in DCIS as compared to LCIS, justifies the current practice of invasive treatments.

Specific estimates for overdiagnosis rate by grade will become increasingly important. These estimates may change when the treatment for DCIS can be even more customised according to grade.(36) To our knowledge a trial to compare treatment of DCIS to active surveillance is planned.(37)

LIMITATIONS OF THE STUDY

We did not review grading or examined interobserver variation between pathologists, because this was beyond the scope of our study. PALGA and the Dutch association of pathologists will be conducting a study to evaluate the interobserver variation in the near future. We believe our study to be a proper representation of the current Dutch situation. There is no reason to suspect that DCIS not detected by mass-screening represents a different patient group than DCIS detected by mass-screening, and for that reason for both groups the same dilemma with regard to a possible interobserver variation exists.

Assumptions on behaviour of DCIS were done on older studies. Advances have been made in the evaluation of biopsies. Currently more sampling is done and pathologists are more aware of the possible findings in DCIS, this could influence the assumptions on behaviour of DCIS if the studies on which they are based were repeated now.

CONCLUSION

DCIS grade is almost equally distributed across the screened population in the breast cancer screening program and the population not subjected to/participating in mass-screening.

DCIS has been divided into three grades, each constituting a unique entity with its own natural history. We found that the distribution of these grades is not dependent on mass-screening status, but is dependent on age. When taking the different grades into account, overdiagnosis rates of breast cancer in mass-screening are 60% for low-grade DCIS and 45% for high-grade DCIS from a population perspective, and 61% and 45% respectively from an individual perspective. When taking the younger ages and high-grade into account overdiagnosis rate from an individual perspective is 21-29%.

These figures underline the necessity of large randomised trials for watchful waiting in low-grade DCIS, whether these are detected in a mass-screening program or not.

REFERENCES

1. Jones JL. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. Breast Cancer Res. 2006;8(2):204.
2. Registry TNC. Cijfersoverkanker [Webpage]. 2013 [cited 2013 17-09-2013]. Available from: http://www.cijfersoverkanker.nl/selecties/dataset_3/img523841be3dd9b.
3. Hofvind S, Lee CI, Elmore JG. Stage-specific breast cancer incidence rates among participants and non-participants of a population-based mammographic screening program. Breast Cancer Res Treat. 2012;135(1):291-9.
4. 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(1):111-21.
5. 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(12):1746-54.
6. Evans AJ, Pinder SE, Ellis IO, Wilson AR. Screen detected ductal carcinoma in situ (DCIS): overdiagnosis or an obligate precursor of invasive disease? J Med Screen. 2001;8(3):149-51.
7. Douglas-Jones AG, Gupta SK, Attanoos RL, Morgan JM, Mansel RE. A critical appraisal of six modern classifications of ductal carcinoma in situ of the breast (DCIS): correlation with grade of associated invasive carcinoma. Histopathology. 1996;29(5):397-409.
8. Moulis S, Sgroi DC. Re-evaluating early breast neoplasia. Breast Cancer Res. 2008;10(1):302.
9. Vos CB, ter Haar NT, Rosenberg C, Peterse JL, Cleton-Jansen AM, Cornelisse CJ, et al. Genetic alterations on chromosome 16 and 17 are important features of ductal carcinoma in situ of the breast and are associated with histologic type. Br J Cancer. 1999;81(8):1410-8.
10. Ellis IO. Intraductal proliferative lesions of the breast: morphology, associated risk and molecular biology. Mod Pathol. 2010;23 Suppl 2:S1-7.
11. Gupta SK, Douglas-Jones AG, Fenn N, Morgan JM, Mansel RE. The clinical behavior of breast carcinoma is probably determined at the preinvasive stage (ductal carcinoma in situ). Cancer. 1997;80(9):1740-5.
12. Tsikitis VL, Chung MA. Biology of ductal carcinoma in situ classification based on biologic potential. Am J Clin Oncol. 2006;29(3):305-10.
13. Early Breast Cancer Trialists' Collaborative G, Correa C, McGale P, Taylor C, Wang Y, Clarke M, et al. Overview of the randomized trials of radiotherapy in ductal carcinoma in situ of the breast. J Natl Cancer Inst Monogr. 2010;2010(41):162-77.
14. de Roos MA, van der Vegt B, de Vries J, Wesseling J, de Bock GH. Pathological and biological differences between screen-detected and interval ductal carcinoma in situ of the breast. Ann Surg Oncol. 2007;14(7):2097-104.

15. Kessar P, Perry N, Vinnicombe SJ, Hussain HK, Carpenter R, Wells CA. How significant is detection of ductal carcinoma in situ in a breast screening programme? Clin Radiol. 2002;57(9):807-14.
16. Meijnen P, Peterse JL, Oldenburg HS, Woerdeman LA, Rutgers EJ. Changing patterns in diagnosis and treatment of ductal carcinoma in situ of the breast. Eur J Surg Oncol. 2005;31(8):833-9.
17. Kerlikowske K, Molinaro AM, Gauthier ML, Berman HK, Waldman F, Bennington J, et al. Biomarker expression and risk of subsequent tumors after initial ductal carcinoma in situ diagnosis. J Natl Cancer Inst. 2010;102(9):627-37.
18. Sorum R, Hofvind S, Skaane P, Haldorsen T. Trends in incidence of ductal carcinoma in situ: the effect of a population-based screening programme. Breast. 2010;19(6):499-505.
19. Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. Radiology. 2012;265(3):707-14.
20. Casparie M, Tiebosch AT, Burger G, Blauwgeers H, van de Pol A, van Krieken JH, et al. Pathology databanking and biobanking in The Netherlands, a central role for PALGA, the nationwide histopathology and cytopathology data network and archive. Cell Oncol. 2007;29(1):19-24.
21. Holland R, Peterse JL, Millis RR, Eusebi V, Faverly D, van de Vijver MJ, et al. Ductal carcinoma in situ: a proposal for a new classification. Semin Diagn Pathol. 1994;11(3):167-80.
22. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985;20(1):79-93.
23. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev. 2010;19(8):1893-907.
24. Burnside ES, Trentham-Dietz A, Kelcz F, Collins J. An Example of Breast Cancer Regression on Imaging. Radiology Case Reports. 2006;1(2):27-37.
25. Dehen R. Regression of Ductal Carcinoma In Situ After Treatment with Acupuncture. J Altern Complement Med. 2013.
26. NETB. Interim report 2011. Main results 2008-2009 breast cancer screening programme in the Netherlands. Report. Rotterdam/ Nijmegen: ErasmusMC/ UMC St Radboud, Department of Public Health/ Department of Epidemiology BaH; 2011.
27. 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(11):2205-40.
28. de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med. 2011;53(3):134-40.

29. Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, Schnitt SJ. Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. Cancer. 2005;103(9):1778-84.
30. Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer. 2005;103(12):2481-4.
31. Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ. JAMA Oncol. 2015;1(7):888-96.
32. Shin HJ, Kim HH, Kim SM, Kwon GY, Gong G, Cho OK. Screening-detected and symptomatic ductal carcinoma in situ: differences in the sonographic and pathologic features. AJR Am J Roentgenol. 2008;190(2):516-25.
33. Weigel S, Hense HW, Heidrich J, Berkemeyer S, Heindel W, Heidinger O. Digital Mammography Screening: Does Age Influence the Detection Rates of Low-, Intermediate-, and High-Grade Ductal Carcinoma in Situ? Radiology. 2015:150322.
34. Millis RR, Ryder K, Fentiman IS. Ductal in situ component and prognosis in invasive mammary carcinoma. Breast Cancer Res Treat. 2004;84(2):197-8.
35. Omer ZB, Hwang ES, Esserman LJ, Howe R, Ozanne EM. Impact of Ductal Carcinoma In Situ Terminology on Patient Treatment Preferences. JAMA Intern Med. 2013.
36. Peres J. DCIS test helps filter at-risk patients. J Natl Cancer Inst. 2012;104(24):1853-5.
37. Elshof LE, Tryfonidis K, Slaets L, van Leeuwen-Stok AE, Skinner VP, Dif N, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. Eur J Cancer. 2015;51(12):1497-510.

CHAPTER 5:

BREAST CANCER INCIDENCE TRENDS IN NORWAY AND ESTIMATES OF OVERDIAGNOSIS.

Paula van Luijt

Eveline Heijnsdijk

Nicolien van Ravesteyn

Solveig Hofvind

Harry de Koning.

J Med Screen. 2016. In press.

ABSTRACT

The incidence of breast cancer in Norway has fluctuated in the last three decades. This is partly explained by the use of hormone replacement therapy (HRT) and mammography screening, overdiagnosis has been suggested as well. The Norwegian Breast Cancer Screening Programme (NBCSP) was gradually implemented between 1996 and 2005. We calibrated our microsimulation model to Norwegian Cancer Registration data. The model takes mammography use, in the NBCSP and outside the programme, and HRT use into account. We obtained a proper fit of breast cancer incidence in recent years, when assuming an increase in the background risk for breast cancer and estimated overdiagnosis. We estimated a 2% overdiagnosis rate as a fraction of all cancers diagnosed in women aged 50-100, and a 3% overdiagnosis rate as a fraction of all cancers diagnosed in women aged 50-70, i.e. the screening age. If all of the increased incidence would be the result of the detection of slow growing tumours, these estimates were 7% and 11%, respectively. Besides mammography and HRT use, additional risk factors have contributed to the sudden increase in breast cancer incidence in Norway. Overdiagnosis estimates due to screening were within the range of international plausible estimates.

INTRODUCTION

Breast cancer incidence in Norway rose sharply in the mid-1990s for women aged 50-69 years old; from 193 per 100,000 women in 1995 to 262 per 100,000 women in 1997 (1). The Norwegian Breast Cancer Screening Programme (NBCSP) was introduced in Norway between 1996 and 2005 (2). The rapid increase in breast cancer incidence for women aged 50-69 started in 1995, prior to the introduction of the NBCSP. Two possible causes for this increase have been described before: First, many women were already screened outside the NBCSP, in the period 1983-1996 (3); second, many women used hormone replacement therapy (HRT) in those days (4-7), which is associated with an increased risk of breast cancer (8, 9). Recent studies on breast cancer incidence in Norway found that the use of mammography and HRT alone is not enough to explain the steep increase in breast cancer incidence in Norway (10, 11).

Mammography screening impacts breast cancer incidence because women are diagnosed earlier, so breast cancer incidence in the screening age increases temporarily, but breast cancer incidence after the screening age will decrease. Also screening will lead to diagnoses of breast cancer in women who would never have had a diagnosis in the situation without screening (12). These women are overdiagnosed. Since it is not known which women are overdiagnosed, all women with a diagnosis of breast cancer will be treated. Thus overdiagnosis is directly related to overtreatment. Overdiagnosis is one of the major harms of screening, and the main topic of controversy regarding mass mammography screening (10, 13-19). Overdiagnosis rates in Norway have been estimated to range from 10-50% (14, 15, 17). This variation is mainly attributable to different methods to estimate overdiagnosis (20) and definitions (12). We use the definitions the UK Independent Review Panel deemed most useful, to allow for maximal comparability (21).

The aim of our study is to assess the trends in breast cancer incidence and overdiagnosis in Norway. First, we evaluated trends in overall breast cancer incidence rate, and ductal carcinoma in situ (DCIS) incidence rate, as provided by the Norwegian Cancer Registry, with joinpoint analysis (22). Second, we calibrated our MIcro-simulation SCreening ANalysis (MISCAN) model to the Norwegian data. Finally we used the model to evaluate the NBCSP and HRT and estimate the amount of overdiagnosis.

MATERIALS AND METHODS

JOINPOINT ANALYSIS

We evaluated trends in invasive breast cancer incidence rate for women aged 50-70 years from 1970 to 2009; DCIS incidence rate for women aged 50-70 years from 1993 to 2009; mammography use from 1995 to 2005; and HRT-use from 1986 to 2008. We performed a joinpoint analysis, using the tool provided on the Surveillance research website of the National Cancer Institute (23, 24). Incidence data were provided by the Cancer Registry of Norway. Figures for mammography use were taken from Lynge et al (3). HRT use was extracted from the Norwegian Prescription Database (25).

THE MISCAN MODEL

The MISCAN-model is a micro-simulation model designed for the evaluation of screening. In MISCAN, individual life histories are simulated, and the consequences of introducing a screening programme are assessed.

The model is a semi-Markov model using Monte Carlo simulation. The model simulates a large number of individual life histories; together these life histories form a population. Year of birth, cause of death (breast cancer death or death from other causes), and time of death are included in the life history. Some individuals have an onset of breast cancer and develop a pre-clinical DCIS. After the onset of disease, the pre-clinical DCIS may regress back to normal, progress to consecutive stages of pre-clinical invasive cancer, or become clinically detected (26). The natural history of breast cancer in the model is shown in Figure 1. Age-dependent dwell times represent the average amount of time between transitions from one stage to the subsequent stage. In each stage there is an age-dependent probability that the cancer will be clinically detected. Once the disease is detected, an age and stage dependent survival is assigned. The model determines the cause of death by taking the first of the two options from the life history: death from breast cancer or death from other causes.

Screening is superimposed on the life histories. Screening changes the course of the life histories by changing the time of diagnosis and the time and cause of death. The model generates two life histories for each individual: one for the situation without screening and one for the situation with screening.

Originally, the model was calibrated to the Dutch situation (12, 26). Survival after clinical diagnoses and the improvement of prognosis after detection by screening were modelled using several international sources (27, 28).

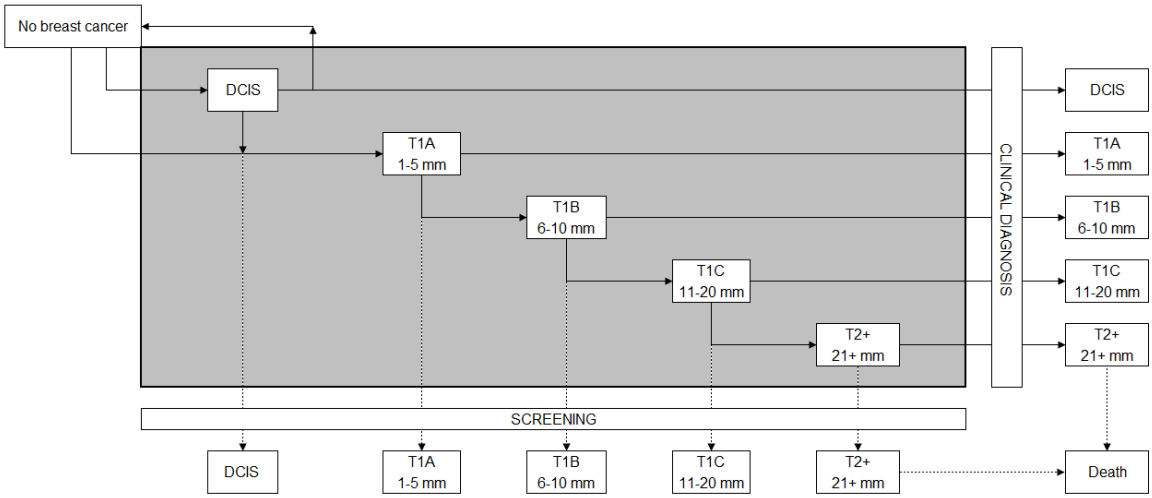


Figure 1. Graphic presentation of the transitions in the MISCAN model. DCIS; ductal carcinoma in situ, T1A; invasive breast cancer with a diameter of 1-5 mm., T1B; invasive breast cancer with a diameter of 6-10 mm.,T1c; invasive breast cancer with a diameter of 11-20mm., T2+; invasive breast cancer with a diameter greater than 21 mm. Every woman has a probability of having an onset of breast cancer. This may be a DCIS or a T1A tumour. From DCIS she may regress back to not having breast cancer, be clinically or screen-detected or progress to T1A. The possibility of regression exists only in the DCIS state.

MODEL ADJUSTMENTS TO THE NORWEGIAN SITUATION

Observed data on breast cancer incidence, stage distribution, detection mode, and attendance rate of the NBCSP were provided by the Cancer Registry Norway.

The Norwegian population was modelled by calibrating the birth table to fit the population composition in 2005 and by replacing the Dutch life table with the Norwegian life table (29). We calibrated the model by fitting the onset rate by age and the stage specific dwell times against the observed incidence rate for the period 1970-1990 (before the wide-spread use of mammography).

Opportunistic screening was included based on the data published by Lynge et al, she found an increased use of mammography from the early 1980's up to approximately 130,000 women examined in 2003.(3) The total number of mammographies in the model was calibrated to the total number of mammographies in the paper . HRT use was implemented as a relative risk for different birth cohorts in individual calendar years (Table 1)(11). The frequency of HRT use was estimated using data from the Norwegian Prescription Database. We used a relative risk of 2.2 of onset of breast cancer for women using HRT, found in previous studies (8, 11).

Table 1 (next page): Relative risk factors used to model increased risk for breast cancer for the whole female population as a result of hormone replacement therapy (HRT). In model 2 an additional relative risk was added to increase the incidence in later years. Note: relative risks in Model 3 are the same as in Model 1.

	Model 1								Model 2							
Birth cohort	1960-1970	1950-1960	1940-1950	1930-1940	1920-1930	1910-1920	1900-1910	1890-1900	1960-1970	1950-1960	1940-1950	1930-1940	1920-1930	1910-1920	1900-1910	1890-1900
Age group	30-46	30-56	37-66	47-76	57-86	67-96	77-99	87-99	30-46	30-56	37-66	47-76	57-86	67-96	77-99	87-99
Year																
1987			1.01	1.01	1.03	1.02	1.02	1.02			1.01	1.01	1.03	1.02	1.02	1.02
1988			1.01	1.07	1.04	1.02	1.02	1.02			1.01	1.07	1.04	1.02	1.02	1.02
1989			1.01	1.08	1.04	1.02	1.03	1.03			1.01	1.08	1.04	1.02	1.03	1.03
1990		1.01	1.09	1.05	1.03	1.03	1.03			1.01	1.09	1.05	1.03	1.03	1.03	
1991		1.02	1.12	1.06	1.03	1.04	1.04			1.02	1.12	1.06	1.03	1.04	1.04	
1992		1.02	1.15	1.08	1.04	1.05	1.05			1.02	1.15	1.08	1.04	1.05	1.05	
1993		1.03	1.20	1.12	1.06	1.06	1.06			1.03	1.20	1.12	1.06	1.06	1.06	
1994		1.04	1.27	1.16	1.07	1.08	1.08			1.04	1.27	1.16	1.07	1.08	1.08	
1995		1.05	1.29	1.19	1.08	1.09	1.08			1.05	1.29	1.19	1.08	1.09	1.08	
1996		1.05	1.31	1.22	1.10	1.10	1.09			1.05	1.31	1.22	1.10	1.10	1.09	
1997		1.06	1.33	1.25	1.11	1.11	1.10			1.75	1.75	1.75	1.75	1.75	1.10	
1998		1.06	1.35	1.28	1.12	1.12	1.11			1.75	1.75	1.75	1.75	1.75	1.11	
1999		1.07	1.35	1.29	1.13	1.13	1.11			1.75	1.75	1.75	1.75	1.75	1.11	
2000	1.06	1.31	1.27	1.12	1.12	1.10			1.75	1.75	1.75	1.75	1.75	1.75		
2001	1.06	1.31	1.27	1.13	1.11	1.10			1.75	1.75	1.75	1.75	1.75	1.75		
2002	1.05	1.27	1.24	1.12	1.10	1.08			1.75	1.75	1.75	1.75	1.75	1.75		
2003	1.04	1.19	1.17	1.09	1.07	1.06			1.75	1.75	1.75	1.75	1.75	1.75		
2004	1.03	1.16	1.14	1.08	1.06	1.05			1.75	1.75	1.75	1.75	1.75	1.75		
2005	1.03	1.13	1.12	1.07	1.05	1.04			1.75	1.75	1.75	1.75	1.75	1.75		
2006	1.02	1.11	1.11	1.06	1.05	1.04			1.75	1.75	1.75	1.75	1.75	1.75		

Incidence rate after the implementation of screening was calibrated, adjusting the parameters age specific hazard (of having an onset of breast cancer), stage specific test sensitivity, the probability of having an onset by age and the incidence of breast cancer. To calibrate these parameters we used data on breast cancer incidence by age after the introduction of screening (1990-2009), disease stage, and detection mode. To reach the observed increase breast cancer incidence, we increased the background incidence for all women under the age of 87 in the years 1997 to 2006 with a relative risk of 1.75. The factor 1.75 is comparable to the cohort effect used by Weedon-Fekjaer in his study.

Mean estimated dwell times are shown in table 2.

Table 2. Dwell time in years by patient age and preclinical disease stage in model 2.

Age	DCIS	T1A	T1B	T1C	T2+
0	2.1	0.0	0.1	0.2	0.3
20	2.1	0.0	0.2	0.4	0.4
30	2.1	0.1	0.2	0.5	0.5
40	2.1	0.1	0.3	0.6	0.7
45	2.1	0.1	0.3	0.7	0.8
50	2.1	0.1	0.4	0.8	0.9
55	2.1	0.1	0.4	1.0	1.1
60	2.1	0.1	0.5	1.2	1.3
65	2.1	0.2	0.7	1.5	1.6
100	2.1	0.2	0.7	1.5	1.6

In sensitivity analysis we also included the outcomes of the model if we would not increase the background risk, and calibrate the dwell time of DCIS to the observed data on breast cancer incidence after the introduction of screening (1990-2009), disease stage, and detection mode. Dwell times in this model were 25.0 years for DCIS, 0.01 year for T1a tumours, 0.1 year for T1b tumours, 0.6 years for T1c tumours and 0.6 years for T2+ tumours. A long dwell time simulates the existence of a large pool of dormant breast cancer. Screening will then mostly detect indolent cancers, and thus create overdiagnosis.

OUTPUT MEASURES

Our analysis was performed on runs that simulated a population of 10 million women for all three models. Rates were calculated per 100,000 women years. We calculated breast cancer incidence with screening, and breast cancer incidence without screening, per 5 year age group for the years 1970-2008.

CALCULATING OVERDIAGNOSIS

In accordance with the UK Independent Review Panel we defined two measures of overdiagnosis: overdiagnosis from a population perspective and overdiagnosis from an individual perspective.

Overdiagnosis from a population perspective = (excess-deficit)/all diagnoses in women exposed to screening aged 50-100.

Overdiagnosis from an individual perspective = (excess-deficit)/all diagnoses in women exposed to screening aged 50-70.

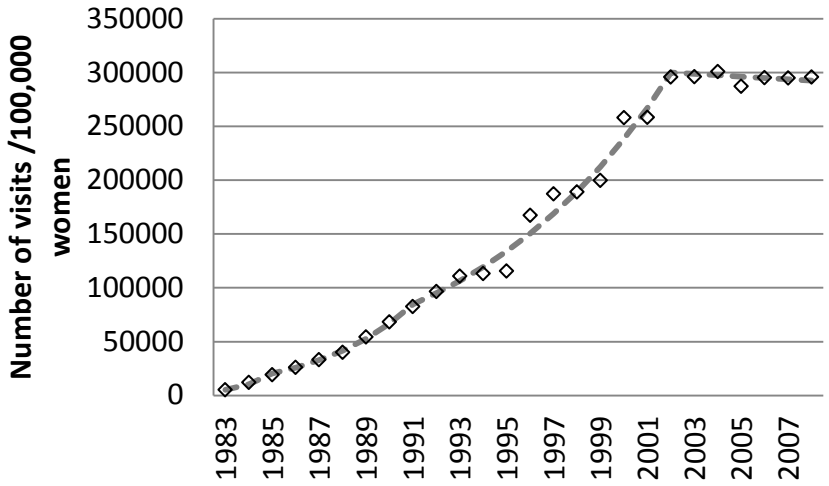
A population, which is exposed to screening, initially has a higher breast cancer incidence compared to a population not exposed to screening (due to earlier diagnoses). These extra diagnoses are called the excess. After the upper age limit of screening is reached, the exposed population initially has a lower breast cancer incidence (breast cancers that were detected in the screening ages, do not occur after screening). The difference between the amount of diagnoses in the entire population (aged 0-100) exposed to screening and the number of diagnoses in the entire population (aged 0-100) not exposed to screening is called the deficit. The total number of overdiagnosed cancers in a population is the difference between excess and deficit.

RESULTS

JOINPOINT ANALYSIS

Mammography use in Norway increased steeply from 1983 to 2002, although subtle, significant changes in annual percentage change were seen in 1985 and 1991. More evidently the use of mammography reached a climax in 2002, after which it gradually declines (Figure 2a). HRT use was increasing steadily between 1990 and 1994, after which the increase became less steep. In 2000 a turning point can be seen, after which the use of HRT rapidly declines (Figure 2b).

A



B

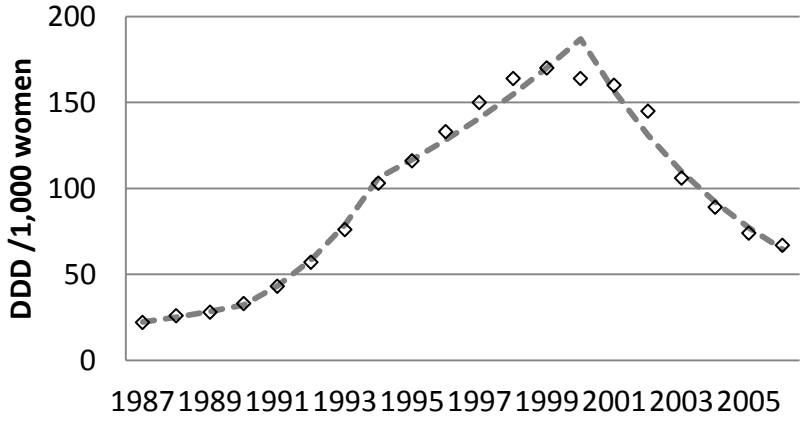
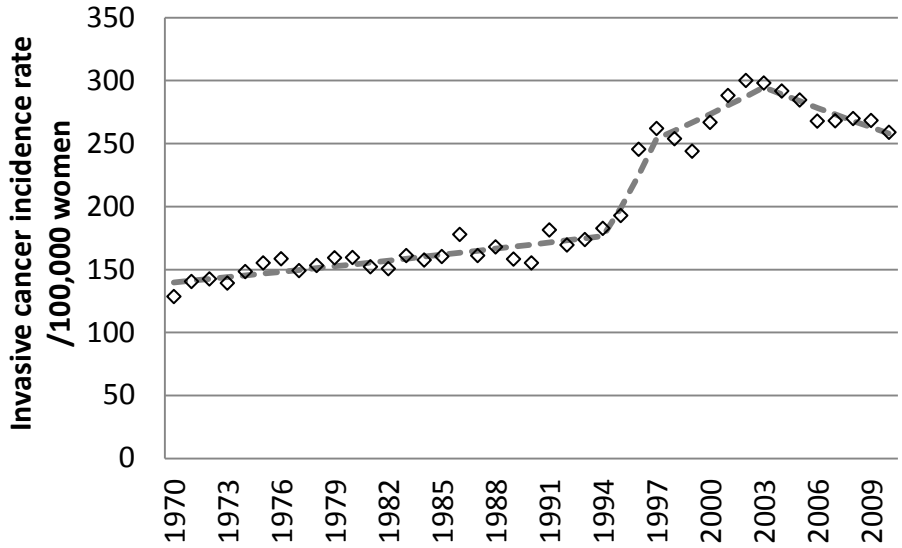


Figure 2. A: Mammography use (both in the NBCSP and opportunistic screening) per 100,000 women of all ages. B: HRT use in defined daily doses (DDD) per 1,000 women of all ages.

Invasive breast cancer was steadily increasing from 1970 to 1995 with 1% per year. From 1995 to 1997 there was a much steeper increase of 13% per year, and after 1997 the increase attenuated to 2.5% per year. The maximum incidence was reached in 2003, after which incidence has been declining with 1.9% per year (Figure 3a). DCIS was increasing rapidly from 1995 to 1997 with 29.5% per year, after which the increase attenuated to 3.3% per year (Figure 3b).

A



B

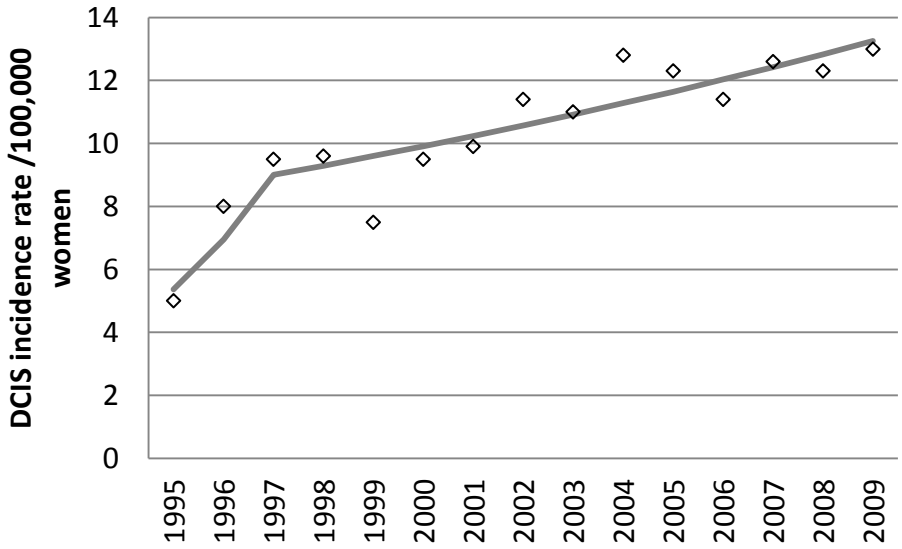


Figure 3. A: Breast cancer incidence rate (invasive cancers only) per 100,000 women aged 50-70. B: DCIS incidence rate per 100,000 women aged 50-70.

MODEL OUTPUT

When assuming an increased background risk the model adequately reproduced the observed breast cancer incidence data. If a stable background risk is assumed, the model underestimates breast cancer incidence in Norway in the years 1997-2009. If we assume a long dwell time for DCIS, incidence rates peak at first introduction of mammography use, but after that incidence rates drop rapidly, and incidence is underestimated in the years 1997-2009 (Figure 4).

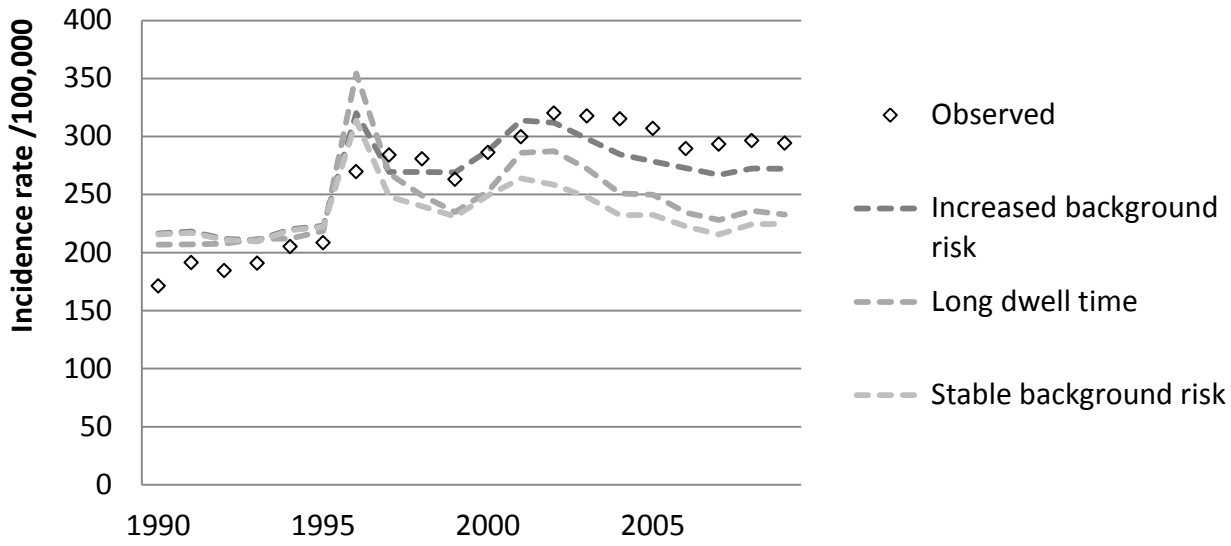


Figure 4. Breast cancer incidence rate/ 100,000 women aged 50-75. The calibrated model to Norwegian demographics, mammography use and HRT use; with increased background risk; model calibrated to Norwegian demographics, mammography use and HRT use, without increased background risk; calibrated model to Norwegian demographics, mammography use and HRT use, without increased background risk, but with very long dwell times for DCIS.

When comparing the output for the situation with screening and without screening, we found an estimated overdiagnosis rate for the years 2014-2023 for all invited women aged 50-100 years (population estimate) of 2% in our best model. For all invited women aged 50-70 (individual estimate) in 2014-2023 this rate was 3%.

The model with long dwell times estimates the highest level of overdiagnosis, because it assumes a large pool of dormant tumours to be detected by screening, which would never have led to breast cancer death or even invasive disease in the absence of screening. In this model estimated overdiagnosis rate for the years 2014-2023 for all invited women aged 50-100 years (population estimate) was 7%. For all invited women aged 50-70 (individual estimate) in 2014-2023 this rate was 11%. Overdiagnosis estimates per model are given in Table 3. Overdiagnosis estimates decrease with a longer follow up, because this allows for the full deficit in the years after screening to manifest.

Table 3 (next page): Outcomes per model. Based on a simulated population of 10,000,000 women. Excess-deficit is the difference of all breast cancer diagnoses (invasive and DCIS) for all ages with screening minus all breast cancer diagnoses (invasive and DCIS) for all ages without screening (diagnoses with screening in ages 50-100 and 50-70). The overdiagnosis estimates follow by dividing excess-deficit by the second and third columns.

	Excess - Deficit per 100,000 women of all ages			Breast cancer diagnosis with screening in ages 50-100 per 100,000 women			Breast cancer diagnosis with screening in ages 50-70 per 100,000 women			Overdiagnosis estimate on a population level			Overdiagnosis estimate on an individual level		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
2000	5.17	7.44	10.00	90.89	103.70	93.59	50.52	58.27	51.38	5.7%	7.2%	10.7%	10.2%	12.8%	19.5%
2001	11.69	14.03	20.27	95.76	110.81	102.57	55.38	65.86	60.64	12.2%	12.7%	19.8%	21.1%	21.3%	33.4%
2002	10.20	12.98	19.50	95.65	112.72	102.44	55.99	66.69	61.79	10.7%	11.5%	19.0%	18.2%	19.5%	31.5%
2003	7.95	10.02	16.20	93.34	110.20	99.09	53.26	64.34	59.19	8.5%	9.1%	16.3%	14.9%	15.6%	27.4%
2004	5.72	7.36	12.67	90.07	109.05	95.97	51.66	63.12	56.02	6.4%	6.7%	13.2%	11.1%	11.7%	22.6%
2005	4.65	5.08	10.85	89.85	107.38	94.61	52.36	62.26	55.86	5.2%	4.7%	11.5%	8.9%	8.2%	19.4%
2006	2.87	3.17	7.94	87.33	106.07	90.53	50.26	60.97	52.94	3.3%	3.0%	8.8%	5.7%	5.2%	15.0%
2007	1.23	2.34	7.68	84.34	103.48	88.44	48.73	59.39	51.55	1.5%	2.3%	8.7%	2.5%	3.9%	14.9%
2008	3.23	3.23	8.42	85.98	104.11	89.68	50.84	61.19	53.24	3.8%	3.1%	9.4%	6.3%	5.3%	15.8%
2009	2.96	3.62	6.81	85.87	103.66	88.15	50.49	61.03	52.54	3.5%	3.5%	7.7%	5.9%	5.9%	13.0%
2010	3.41	3.09	7.11	85.29	103.23	88.55	51.52	61.66	54.25	4.0%	3.0%	8.0%	6.6%	5.0%	13.1%
2011	1.39	2.90	7.24	87.67	106.91	88.32	53.40	65.07	54.47	1.6%	2.7%	8.2%	2.6%	4.5%	13.3%
2012	3.27	2.94	7.37	87.74	104.84	90.30	53.80	63.61	55.87	3.7%	2.8%	8.2%	6.1%	4.6%	13.2%
2013	1.97	1.99	5.80	89.00	105.55	91.39	54.97	65.44	56.30	2.2%	1.9%	6.3%	3.6%	3.0%	10.3%
2014	0.66	2.80	6.30	89.36	107.05	93.25	53.84	65.24	56.79	0.7%	2.6%	6.8%	1.2%	4.3%	11.1%
2015	2.21	1.87	5.95	92.11	111.10	95.73	56.54	67.93	58.80	2.4%	1.7%	6.2%	3.9%	2.7%	10.1%
2016	1.80	2.61	7.67	96.26	114.15	97.49	58.96	70.20	59.53	1.9%	2.3%	7.9%	3.1%	3.7%	12.9%
2017	3.32	3.88	6.77	96.45	116.40	99.44	59.39	70.27	61.78	3.4%	3.3%	6.8%	5.6%	5.5%	11.0%
2018	2.39	4.02	8.22	98.02	118.91	100.83	60.56	72.99	62.06	2.4%	3.4%	8.2%	3.9%	5.5%	13.3%
2019	1.35	0.55	5.88	100.36	120.79	103.21	59.81	72.42	62.80	1.3%	0.5%	5.7%	2.2%	0.8%	9.4%
2020	1.53	0.55	6.15	101.58	125.38	105.74	61.97	75.53	64.38	1.5%	0.4%	5.8%	2.5%	0.7%	9.6%
2021	2.28	3.73	7.70	104.89	125.02	107.24	63.55	77.10	66.08	2.2%	3.0%	7.2%	3.6%	4.8%	11.7%
2022	3.07	2.47	7.07	108.83	130.04	108.68	65.18	77.02	65.82	2.8%	1.9%	6.5%	4.7%	3.2%	10.7%
2023	1.23	2.37	5.33	106.60	131.56	112.41	63.45	79.41	66.37	1.2%	1.8%	4.7%	1.9%	3.0%	8.0%
2014-2023	19.8	24.85	67.04	994.46	1200.42	1024.01	603.26	728.1005	624.42	2%	2%	7%	3%	3%	11%

DISCUSSION

In this study, we confirmed that the steep increase in breast cancer incidence cannot be explained by mammography use and HRT use alone. We explained the additional increase with a relative risk of 1.75 for women aged 87 and younger in the years 1997-2006.

The increase in breast cancer incidence in 1994 occurred three years after the introduction of mammography, and was therefore probably the cumulative result of increased HRT use and increasing mammography use. The relative decline in 1997 coincided with a relative decline of DCIS. Since this is a condition strongly associated with mammography, it seems likely that this change is the result of the fact that by then many women have had a prevalence screening; the increase in incidence as a result of detection of prevalent tumours was decreasing. The rapid decrease in 2003 coincided with the stabilization of mammography use, and followed three years after the rapid decrease in the use of HRT, and thus was most likely a combined result of these factors.

In order to fit our model to the data we increased the background risk, similar to the cohort effect used by Weedon-Fekjaer(11). This is also in line with the findings of Duffy et al. who found that not all increase in incidence could be explained by extrapolating an age-specific period effect model or an age-adjusted common period effect based on data prior to the introduction to screening, to the years after the introduction of screening, and that this excess incidence could not be attributed to screening alone (10). The problem with calibrating the relative risk to the empirical incidence is that we deny the possibility that all of this increase could be due to overdiagnosis. Therefore we included the analysis with the model with long dwell times, which aims to attribute the excess incidence to overdiagnosis. This model however does not fit the data as well, and the relative risk is externally validated by the findings of Weedon-Fekjaer, who found a similar cohort effect.(11) They found an increase in risk from 0.74 to 1.25, a factor 1.7, which correlates nicely to our RR of 1.75.

A possible explanation for the unaccounted increase in incidence may be that Norwegian women may have a larger additional risk of breast cancer due to HRT use. This could be the result of differences in doses, combinations of hormones in preparations, and duration of use (8, 11). We used a relative risk of 2.2, which is among the highest in literature, and still could not fully explain the increase in incidence.

Miscellaneous factors may be responsible. A possible association has been suggested between breast cancer and age at menarche, age at first birth, physical activity, alcohol, and general change in life style factors. Number of children may play a role, however fertility rates in Norway have been stable and the rate of childlessness at the age of 45 is relatively low (30). Also refraining from breast feeding has been suggested, however lactation is very common in Norway(31). Tall women may also be at greater risk, Norwegians are among the tallest people of the world (32).

In light of this discussion it would be interesting to look at the histopathological characteristics of breast cancers detected by screening, vs breast cancers detected outside the screening programme. A previous study showed a more favourable stage-distribution among participants in the NBCSP. This indicates early detection on one hand, but may also be the result of the detection of slow growing tumours with low malignant potential (33).

If we do not assume an increased background risk, our model is comparable to the study of Falk et al (14). They showed that the estimated incidence rate ratio due to screening is approximately 1.86 for prevalence mammography (women attending screening have an incidence rate of 1.86 times that of women who do not attend screening). The incidence rate ratio was 1.46-1.69 for incidence mammography and 0.60-0.92 for women aged 70-80 (after the end of screening has been reached). We compared the impact of HRT and mammography, based on literature, to the background incidence, which is an extrapolation of trends in the years 1970-1990. The overdiagnosis estimate from this model was low for the period 2014-2023. The estimates in the early 2000s corresponded to the estimate made by Falk et al. (11-21%, results not shown).

When we attributed all excess increase in incidence to the detection of prevalent, slow growing tumours (assuming no regression of invasive disease), this provided a poorer fit to the incidence data, but it also provided the highest estimate of overdiagnosis rate. The results stated in the report of the Research Council of Norway indicated their overdiagnosis estimates to be within this range. (34) (34) (34) (34)

Our overdiagnosis estimates in the early 2000s were in line with earlier estimates published for the same time periods (14, 17). These estimates were between 10 and 25%. Our overdiagnosis estimates for later years were much lower and emphasised the need for sufficient follow-up before calculating overdiagnosis, to allow the complete occurrence of the deficit (12).

CONCLUSION

The increase in breast cancer incidence in Norway cannot be fully explained by mammography screening and HRT use. We calculated models to estimate the impact of an increased back ground incidence and a large impact of screening. We estimate overdiagnosis rates at 2-3%, with a maximum estimation of 7-11% for the period 2014-2023. It will be very interesting to see what will happen to breast cancer incidence in the upcoming years, as we move further from the massive use of HRT and have a prolonged steady state screening programme in place.

REFERENCES

1. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol. 2010;49(5):725-36.
2. Hofvind S, Wang H, Thoresen S. Do the results of the process indicators in the Norwegian Breast Cancer Screening Program predict future mortality reduction from breast cancer? Acta Oncol. 2004;43(5):467-73.
3. Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, et al. Mammography activity in Norway 1983 to 2008. Acta Oncol. 2011;50(7):1062-7.
4. Kerlikowske K, Cook AJ, Buist DS, Cummings SR, Vachon C, Vacek P, et al. Breast cancer risk by breast density, menopause, and postmenopausal hormone therapy use. J Clin Oncol. 2010;28(24):3830-7.
5. Sogaard AJ, Tollan A, Berntsen GK, Fonnebo V, Magnus JH. Hormone replacement therapy: knowledge, attitudes, self-reported use - and sales figures in Nordic women. Maturitas. 2000;35(3):201-14.
6. Bakken K, Eggen AE, Lund E. Hormone replacement therapy in Norwegian women, 1996-1997. Maturitas. 2001;40(2):131-41.
7. Hemminki E, Kyyronen P, Pukkala E. Postmenopausal hormone drugs and breast and colon cancer: Nordic countries 1995-2005. Maturitas. 2008;61(4):299-304.
8. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet. 2003;362(9382):419-27.
9. Bakken K, Alsaker E, Eggen AE, Lund E. Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. Int J Cancer. 2004;112(1):130-4.
10. Duffy SW, Michalopoulos D, Sebuodegard S, Hofvind S. Trends in aggregate cancer incidence rates in relation to screening and possible overdiagnosis: A word of caution. J Med Screen. 2014;21(1):24-9.
11. 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.
12. 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(1):111-21.
13. Hofvind S, Ursin G, Tretli S, Sebuodegard S, Moller B. Breast cancer mortality in participants of the Norwegian Breast Cancer Screening Program. Cancer. 2013;119(17):3106-12.
14. Falk RS, Hofvind S, Skaane P, Haldorsen T. Overdiagnosis among women attending a population-based mammography screening program. International journal of cancer Journal international du cancer. 2013;133(3):705-12.

15. Zahl PH, Maehlen J. Overdiagnosis of breast cancer after 14 years of mammography screening. Tidsskr Nor Laegeforen. 2012;132(4):414-7.
16. Zahl PH, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med. 2008;168(21):2311-6.
17. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Annals of internal medicine. 2012;156(7):491-9.
18. Weedon-Fekjaer H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study. BMJ. 2014;348:g3701.
19. Sorum R, Hofvind S, Skaane P, Haldorsen T. Trends in incidence of ductal carcinoma in situ: the effect of a population-based screening programme. Breast. 2010;19(6):499-505.
20. Puliti D, Miccinesi G, Paci E. Overdiagnosis in breast cancer: design and methods of estimation in observational studies. Prev Med. 2011;53(3):131-3.
21. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. British journal of cancer. 2013;108(11):2205-40.
22. Jones RH, Dey I. Determining one or more change points. Chem Phys Lipids. 1995;76(1):1-6.
23. Joinpoint Regression Program: National Cancer Institute

[updated 29 Dec 2014; cited 2014]. Available from: http://surveillance.cancer.gov/joinpoint/.
24. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000;19(3):335-51.
25. FolkeHelseInstituttet. Norwegian Prescription Database http://wwwnorpdno/. 2004;2013.
26. de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Preventive medicine. 2011;53(3):134-40.
27. de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. Journal of the National Cancer Institute. 1995;87(16):1217-23.
28. Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am. 2000;38(4):625-51.
29. StatisticsNorway. wwwssbno

2014.
30. Lappegård T. New fertility trends in Norway. Demographic Research. 2000;2:3.
31. Grovslien AH, Gronn M. Donor milk banking and breastfeeding in Norway. J Hum Lact. 2009;25(2):206-10.

32. Hatton TJ, Bray BE. Long run trends in the heights of European men, 19th-20th centuries. Econ Hum Biol. 2010;8(3):405-13.
33. Hofvind S, Lee CI, Elmore JG. Stage-specific breast cancer incidence rates among participants and non-participants of a population-based mammographic screening program. Breast Cancer Res Treat. 2012;135(1):291-9.
34. Research-based evaluation of the Norwegian Breast Cancer Screening Program. The Research Council of Norway, May, 2015. Report No.

CHAPTER 6:

COST-EFFECTIVENESS OF THE NORWEGIAN BREAST CANCER SCREENING PROGRAM.

Paula van Luijt

Eveline Heijnsdijk

Harry de Koning.

2016. Submitted.

ABSTRACT

The Norwegian Breast Cancer Screening Programme (NBCSP) has a nation-wide coverage since 2005. All women aged 50-69 years are invited biennially for mammography screening. We evaluated breast cancer mortality reduction and performed a cost-effectiveness analysis, using our microsimulation model, calibrated to most recent data. The microsimulation model allows for the comparison of mortality and costs between a (hypothetical) situation without screening and a situation with screening. Breast cancer incidence in Norway had a steep increase in the early 1990s. We calibrated the model to simulate this increase and included recent costs for screening, diagnosis and treatment of breast cancer and travel and productivity loss.We estimate a 16% breast cancer mortality reduction for a cohort of women, invited to screening, followed over their complete lifetime. Cost-effectiveness is estimated at NOK 112,162 per QALY gained, when taking only direct medical costs into account (the cost of the buses, examinations, and invitations). We used a 3.5% annual discount rate. Cost-effectiveness estimates are substantially below the threshold of NOK 1,926,366 as recommended by the WHO guidelines. For the Norwegian population, which has been gradually exposed to screening, breast cancer mortality reduction for women exposed to screening is increasing and is estimated to rise to approximately 30% in 2020 for women aged 55-80 years. The NBCSP is a highly cost-effective measure to reduce breast cancer specific mortality. We estimate a breast cancer specific mortality reduction of 16 to 30%, at the cost of 112,162 NOK per QALY gained.

INTRODUCTION

The Norwegian Breast Cancer Screening Program (NBCSP) was initiated in 1996 and gradually expanded to provide nation-wide coverage in 2005.(1) All women aged 50-69 years, based on the Central Population Register, are invited every other year for mammography screening. Breast cancer incidence increased in the early 1990s, as a result of screening, the use of hormone replacement therapy and increasing back ground risk.(2, 3)

The cost-effectiveness of the Norwegian screening program and the achievable breast cancer mortality reduction was reported by Norum and Wang et al. in 1999 and 2001 respectively.(4, 5) Norum used an estimated 30% reduction in breast cancer deaths at the cost of £8,561 (12,971 USD at the exchange rate in 2000) per life year saved. Wang calculated several scenarios at different positive predictive values of the program. At a maximum estimated breast cancer mortality reduction of 20%, the costs were 5,622 USD per life year saved, and at a maximum estimated mortality reduction of 40% the costs were 2,813 USD per life year saved.

The use of mass screening programs for breast cancer has been under debate since 2001, when the Cochrane collaboration first published a systematic review on the topic.(6) Kalager investigated the mortality reduction, by comparing historical groups, based on the staggered implementation of the NBCSP. They found a mortality reduction of only 10%. (7, 8) The focus of the debate is on the estimated breast cancer mortality reduction and the estimated overdiagnosis rate. These estimates vary greatly depending on the analysis chosen to correct for lead time and historical bias, the minimal amount of follow up to allow for full dwell time, and the definition of denominator and numerator(9-11). Historical bias occurs when incidence rates rise, independent of screening, due to increases in background incidence. Recently the Research Council of Norway reviewed the literature and had several analyses conducted on breast cancer mortality reduction. They found estimates between 7 and 30%. They concluded based on the quality of the studies that a mortality reduction of 20-30% is a reasonable estimate.(12)

In 2008, the Research Council of Norway set up a project to evaluate the NBCSP. “The objective of the research-based evaluation was to investigate whether the NBCSP fulfils its intentions and purpose to estimate.”(12) This study was done as a part of this project. The council concluded that “The estimates indicate that the Norwegian program performs on average at the level that could be expected from the majority of previous reviews of the mammography screening trials.” This is in line with the conclusions of the International Agency for Research on Cancer, as published in their handbook.(13)

The aim of the study is to estimate the breast cancer mortality reduction due to screening and to calculate cost-effectiveness in the NBCSP. We established trends in breast cancer mortality, using joinpoint analysis, and simulated the NBCSP using microsimulation to estimate expected mortality reduction as a result of screening, and calculate cost-effectiveness by QALY gained for women invited to screening, using direct and indirect costs of screening and treatment.

MATERIALS AND METHODS

JOINPOINT ANALYSIS

Data on breast cancer mortality were obtained from Nordcan.(14) Breast cancer mortality was given as a crude rate and calculated by year. We evaluated breast cancer mortality trends with joinpoint analysis, with the tool provided on the Surveillance research website of the National Cancer Institute.(15) We analysed breast cancer mortality in women aged 55-80 and 0-100 years in the years 1984-2011, and allowed for a maximal of 5 joinpoints.

MODEL DESCRIPTION

The MIcro-simulation SCreening ANalysis (MISCAN) model simulates a female population by simulating individual life histories from birth to death.(16-18) The age-composition of the study population was determined by calibrating the model with the birth table and life table from Statistics Norway from 2005. Each woman has a probability of an onset of breast cancer, based on incidence rate, which we obtained from the Cancer Registry of Norway. The first transition from disease free to preclinical disease is dependent on the onset. Each preclinical disease stage has two possibilities: the disease can be clinically detected or the disease can have a transition from the first preclinical disease state to the next preclinical disease state. Preclinical ductal carcinoma in situ (DCIS) can also regress back to normal. Progression of breast cancer is modelled in a pseudo Markov transition model (Figure 1). In the model any woman can only develop breast cancer once in her life history. Treatment is implemented in the model based on data from the Dutch screening organisations and the Eindhoven Cancer Registry.(19, 20) Early detection has an improved prognosis and therefore reduced mortality. Breast cancer mortality is stage-dependent and based on the data from the Swedish breast cancer screening trials.(21, 22) Screening is superimposed on the natural history. Data from the Cancer Registry of Norway on coverage and attendance, by age (we used five year age groups), year (1990-2010), and stage for the whole country were used to model screening attendance by age and year.

We performed two runs; one population run, based on the Norwegian composition of the population in 2005 (used to estimate current and future breast cancer mortality reduction in the population), and a cohort run, based on an imaginary cohort of 10,000,000 women all born in 1955, with complete follow-up to 2055. The model assumes all women die at the age of 100 years at the latest. In the cohort run we only included the NBCSP, and not opportunistic screening for the cost-effectiveness analysis. All output used in the cost-effectiveness analysis is from the cohort run.

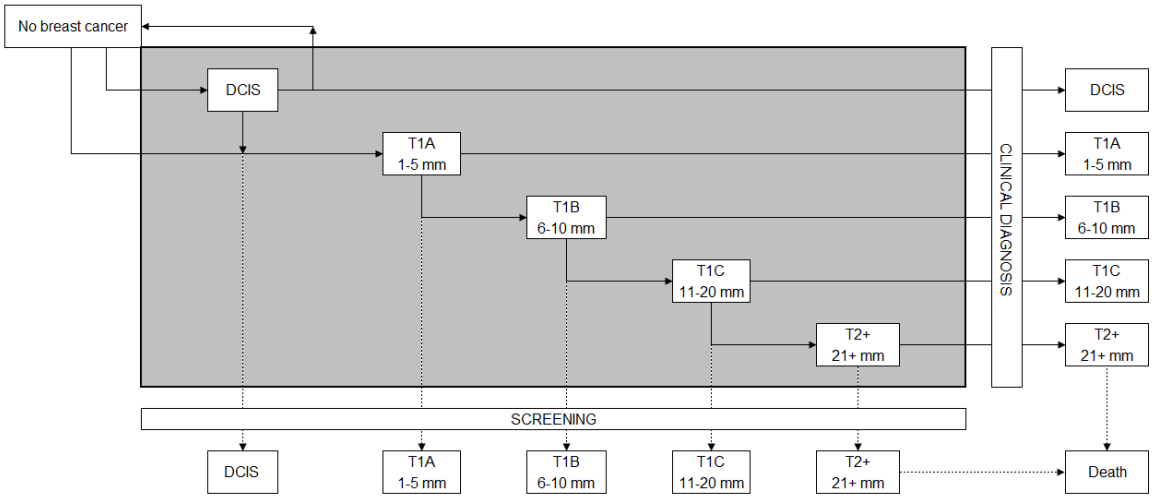


Figure 1. Graphic presentation of the transitions in the MISCAN model. DCIS; ductal carcinoma in situ, T1A; invasive breast cancer with a diameter of 1-5 mm., T1B; invasive breast cancer with a diameter of 6-10 mm.,T1c; invasive breast cancer with a diameter of 11-20mm., T2+; invasive breast cancer with a diameter greater than 21 mm. Every woman has a chance of having an onset of breast cancer. This may be a DCIS or a T1A tumour. From DCIS she may regress back to not having breast cancer, be clinically or screen-detected or progress to T1A. The possibility of regression exists only in the DCIS state.

We performed two runs; one population run, based on the Norwegian composition of the population in 2005 (used to estimate current and future breast cancer mortality reduction in the population), and a cohort run, based on an imaginary cohort of 10,000,000 women all born in 1955, with complete follow-up to 2055. The model assumes all women die at the age of 100 years at the latest. In the cohort run we only included the NBCSP, and not opportunistic screening for the cost-effectiveness analysis. All output used in the cost-effectiveness analysis is from the cohort run.

ADDITIONAL DATA

Screening outside the NBCSP was included based on the data published by Lynge et al, which took place between 1983 and 2008.(23) HRT-use was implemented as an additional risk. The frequency of HRT use was estimated using data from the Norwegian Prescription Database by ten year age group from 1987 to 2006, from the Norwegian Institute of Public Health.(24) We used a relative risk of 2.2 to have an onset of breast cancer for HRT-users, the relative risk found in previous studies. (3)

Observed incidence rates were higher than the estimates made by the model based on mammography use and HRT-use alone. To allow for the steep increase in breast cancer incidence in Norway we added an additional risk factor of 1.75 in the years 1997-2010 for all

women aged 30-100. The additional risk factor creates more onsets and thereby increases breast cancer incidence. This is similar to the effect of HRT-use, but stronger on a population level. A detailed description of the model can be found in a previous paper, which is in press. (25)

BREAST CANCER MORTALITY REDUCTION

We compared the observed breast cancer mortality to the estimated mortality rate in the model (Figure 2). To calculate mortality reduction we compared mortality rate in the situation with screening to the mortality rate without screening. We estimated the expected breast cancer mortality reduction from 2014 to 2034.

The survival rate after treatment was modelled on international sources and the Swedish randomized controlled trials. Details are described in the article of de Gelder et al. (20)

COSTS AND EFFECTS

Direct medical costs including costs of screening, diagnostics and treatment were used in this analysis. The direct non-medical costs of screening are the costs for travel for screening and follow-up examination, and the indirect costs are the costs as a result of productivity loss (societal perspective). All of these costs were obtained from the University of Oslo.(26)

The cost of screening was calculated by multiplying the number of visits with the average screening costs. The average costs per woman attending screening are 812 NOK (83.72 Euro), when only taking direct medical costs into account, or 1,262 NOK (130.11 Euro) when also taking direct non-medical costs and indirect costs (productivity loss) into account. (26) The number of false positives was calculated using the positive predictive value of screening. This is 12.5% in Norway in the years 1996-2005; for every screen detected cancer, 8 women were evaluated after a recall.(27)

To calculate the cost of a diagnosis for a woman who has not been screened we calculated the number of women examined per diagnosis. Positive predictive value of a diagnostic mammography is 59%.

(28) We used the number of false positives as a measure for the number of women undergoing a diagnostic procedure following a diagnostic mammography. The detection rate for symptomatic women is 10.3/1,000 examinations.(29)

The costs for treatment according to stage were provided by the University of Oslo.(30) The costs are shown in Table 1. The initial costs are the costs in the first 6 months since diagnosis. If a woman survives she will have continuous care up to 10.5 years after diagnosis. If she dies of breast cancer she will have six months of terminal costs. If she dies of other causes within 10 years since diagnosis, she will have continuous care up to her death.Treatment costs were given by disease stage: DCIS, TNM I, TNM II, TNM III, and TNM

IV; and in three time frames: initial treatment, for the first six months since a diagnosis; continuous care, from the seventh month up to ten years and six months since diagnosis; and terminal care, the last six months prior to death from breast cancer. The probability of receiving a certain treatment was based on data from the Norwegian Cancer Register.

Table 1. Estimated costs for treatment by disease stage and period.(30)

Initial costs for first 12 months following diagnosis:

	DCIS	TNM1	TNM2	TNM3	TNM4
Mean:	70,642.63	106,868.50	214,542.19	263,548.04	247,895.15
95% CI:	65,000-77,000	102,000-111,000	205,000-224,000	232,000-294,000	200,000-301,000
SE:	3	2.3	4.8	16	26

Continuous care per two months following the first 12 months after diagnosis:

	DCIS	TNM1	TNM2	TNM3	TNM4
Mean:	1,034.52	1,643.52	3,125.05	4,347.90	9,111.90
95% CI:	837-1,240	1,481-1,814	2,862-3,389	3,022-5,849	6,945-11,539
SE:	102	84	133	721	1.158

Terminal costs (last 6 months before death):

	DCIS	TNM1	TNM2	TNM3	TNM4
Mean:	174,504.20	133,712.30	168,846.80	138,925.30	182,511.20
95% CI:	124,000-234,000	106,000-161,000	153,000-186,000	97,000-181,000	142,000-226,000
SE:	28	14	9	21	21

Women who died of breast cancer within six months since their diagnosis were assumed to only have terminal costs. Women who died of breast cancer between 6 months and ten years since their diagnosis were assumed to have initial costs, some continuous care costs, and terminal costs. Women who died of breast cancer after 10.5 years were assumed to have initial costs, all continuous care costs, and terminal costs.

We assumed that women who died from other causes did not receive terminal breast cancer care. If these women died within the first six months since their diagnosis, we calculated the cost of treatment for the time they were alive since their diagnosis. The treatment costs for a woman who died from other causes after six months since diagnosis was calculated by six months of initial care and continuous care for the time they were alive in the period of continuous care.

Because we had costs per timeframe, we used model output on cause of death, number of deaths in the first six months after diagnosis, between six months and 10.5 years after diagnosis, and after 10.5 years after diagnosis. We also used number of diagnosis per stage, the lifeyears in each timeframe and number of visits.

The effect of screening was estimated by calculating the life years gained (LYG). We calculated quality adjusted life years (QALYs) by adjusting life years with utilities, such as described by Haes et al.(31) We adjusted for the screening, diagnostic phase, therapy, disease free survival, terminal illness and palliative care. Based on the data from the Norwegian Cancer Registry we calculated the probability of a certain event by disease stage, and multiplied the number of life years in every disease stage with the probability and the utility. The summarized utility loss and quality adjustment are in table 2 (Table 2). From left to right the columns list the utility loss based on the publication of Haes, the number of events per 100,000 women years in the situation without screening, and in the situation with screening, the difference between these two situations, the average duration associated with an event, the resulting qualitity of life adjustment, the total of life years gained in the situation with screening, the total loss of quality of life, and finally the quality of life adjusted life years gained. Costs and effects were calculated for a cohort of 10 million women born in 1955 and followed until death. Both effects and costs were discounted at 3.5% per year to take time preference into account (NICE).(32) The cost-effectiveness ratio (CER), costs per QALY compared to a situation without screening, was calculated.

Table 2. Utilities and quality adjustment. LY: life years, QALY: quality adjusted life year.

Health stage	Utility loss	No screening	Screening	Difference	Duration §	Quality adjustment	LY gained	Quality of life lost	QALY gained
Per 100,000 women aged 50 years in 2005 with complete follow-up:									
Screening	0.01		586,555	586,555	0.0962	67.68			
Diagnostic phase	0.11	14,025	14,157	132	0.0192	1.33			
Initial surgery	0.13	9,913	10,000	87	0.1667	1.92			
Initial radiotherapy	0.20	6,815	6,797	-18	0.1667	-0.60			
Initial chemotherapy	0.28	2,367	2,134	-233	0.5000	-32.93			
Initial hormonal therapy	0.18	4,098	3,760	-338	2.0000	-121.80			
Terminal illness	0.71	6,002	5,611	-391	0.0833	-23.18			
Palliative therapy + chemotherapy	0.47	1,554	1,358	-196	0.3333	-30.70			
Palliative therapy + radiotherapy	0.42	4,156	3,861	-296	0.0833	-10.32			
Palliative therapy + surgical therapy	0.38	5,973	5,582	-391	0.0962	-14.40			
Palliative therapy + hormonal therapy	0.34	2,646	2,346	-300	1.1667	-118.05			
							6,390.00	304.95	6,085.05
In life years:									
Disease free 2m-1y mastectomy	0.16	84	78	-6	0.8333	-0.80			
Disease free 2m-1y breast conserving	0.09	149	139	-10	0.8333	-0.72			
In life years:									
Disease free >1y mastectomy	0.05	49,674	53,309	3,636	1.0000	192.68			
Disease free >1y breast conserving	0.04	90,056	99,927	9,871	1.0000	394.83			

RESULTS

JOINPOINT ANALYSIS

Crude breast cancer mortality rate for women aged 55-80 rose steadily from 1981 to 1995. In 1995 a significant difference in the annual percent change was found, when breast cancer mortality rates began to drop, coinciding with the start of breast cancer screening, but five years after opportunistic screening came up. For the age group 0 to 100 breast cancer mortality was already declining from 1984, in 1994 a significant percentage change was seen, and from 1994 breast cancer mortality decreased more rapidly (Figure 2).

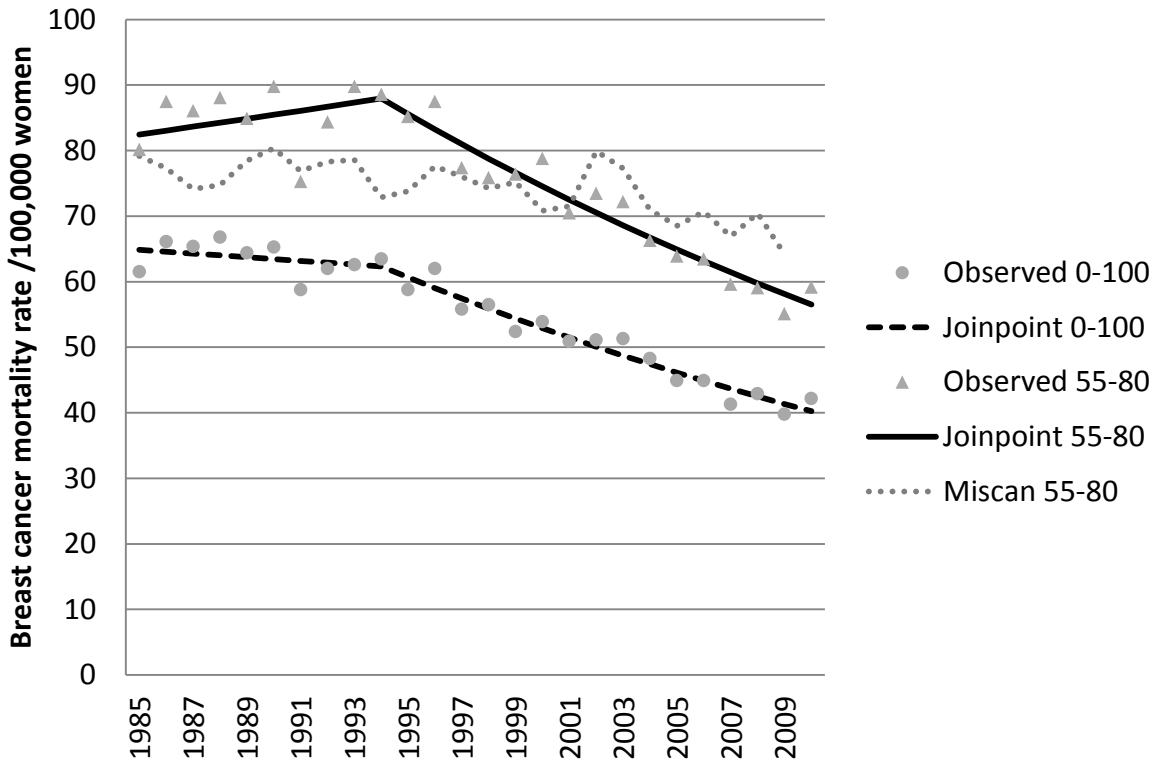


Figure 2. Joinpoint analysis and MISCAN model estimate of breast cancer specific mortality rate per 100,000 women aged 0-100 and 55-80.

MORTALITY

Based on a population run (i.e. a realistic representation of the population of Norway, gradually exposed to screening) of women aged 55-80 in the years 2014-2034. Because we anticipate that mortality will decrease up to 2022 we chose a long follow-up. 2014 was chosen because the program had been nation-wide from 2004, so in 2014 it had been in place for 10 years, and there was a steady situation. Breast cancer mortality reduction was expected to increase up to the year 2022. The reduction was maximally 30%, compared to a situation without screening. The model underestimates the total mortality reduction (Figure 2).

COST-EFFECTIVENESS ANALYSIS

Results are given per 100,000 women (Table 3). Our model showed a life-time breast cancer mortality reduction of 16% for the entire cohort, with a number needed to screen of 1,470 (total number of screens divided by breast cancer deaths prevented). We estimated 399 breast cancer deaths prevented, with approximately 10 deaths prevented annually in the age group 49-100, with a maximum number of 20 deaths prevented in the year 2027 when the cohort is 72 years. The number of QALYs gained was 6,085 /100,000, the number of life-years gained was 6,390, there was a loss of 5% when calculating QALYs from life-years.

The costs per QALY gained are given for two calculations: only including the direct medical costs of screening (NOK 812); and including all direct medical, non-medical and indirect costs of screening (NOK 1,262). Cost effectiveness was NOK 112,162 for only direct medical costs and NOK 189,557 for all costs.

Table 3. Effects, costs and breast cancer mortality reduction per 100,000 women, aged 49 in 2004, with complete follow-up (to 2055), with screening in the NBCSP. QALY: quality adjusted life years. Only the incremental costs/QALY gained are given with a 3.5% annual discount. NOK: Norwegian Krones.

	Without screening	With screening
Screening tests/100,000	-	586,555
Health effects		
Cancer diagnosed	9,968	10,062
Screen detected cancers	-	2,323
False positives	16,945	31,743
Breast cancer deaths	2,425	2,026
QALY gained	-	6,085
Mortality reduction	-	16%
Number needed to screen	-	1,470
Costs (in NOK x1,000)		
Direct medical costs of screening	-	476,283
Total costs of screening	-	740,232
Diagnosis	77,214	119,937
Treatment	2,515,048	2,324,794
Total costs based on direct medical costs of screening	2,592,263	2,921,013
Total costs based on total costs of screening		3,184,963
Incremental costs based on direct medical costs of screening	-	328,750,348
Incremental costs based on total costs of screening		592,700,098
Cost-effectiveness		
Direct medical costs of screening		
Incremental costs/QALY gained, discounted in NOK		112,162
Total costs of screening		
Incremental costs/QALY gained, discounted in NOK		189,557

DISCUSSION

The NBCSP is cost-effective even when taking direct non-medical and indirect costs of screening into account, assuming a threshold for cost-effectiveness of 3 times the gross domestic product (GDP) per capita (WHO guideline).(33) The GDP per capita in Norway was USD 99,636 in 2013 (The World Bank).(34) The exchange rate from USD to NOK on 02/10/2014 was 6.44468. The cost-effectiveness threshold thus is 99,636 x 6.44468 x 3 = 1,926,366 NOK.

The costs of screening in Norway are relatively high compared to other countries; this is probably due to the fact that it is a geographically vast country with a relatively small population. The country has a very high welfare level, which results in a high cost-effectiveness threshold. A recent report in the NEJM advocated the use of a threshold of 100,000 USD (NOK 644,680).(35) Therefore, the NBCSP is also cost-effective when using this threshold. This analysis predicts that breast cancer screening reduces breast cancer mortality at a cost-effective price.

With regards to costs of the program in terms of overdiagnosis, the independent UK review found an estimate of overdiagnosis of 11% acceptable, we found 2-3% overdiagnosis in a separate analysis.(12, 36)

Norum found a lower estimate of cost-effectiveness of NOK 89,325 (calculated from the published £8,561 with the current exchange rate of 10.4339).(4) At the time of their analysis no data on breast cancer mortality reduction was available and they assumed a breast cancer mortality reduction of 30%, accordant with the aim of the Norwegian Mammography Project. The results are difficult to compare, since they discounted by 5% per year and they did not adjust for quality of life.(4)

The cost-effectiveness analysis of Wang estimated much lower costs per life-year gained (NOK 24,167, calculated from the published USD 3,750 with the current exchange rate). These costs did not include treatment costs and did not adjust for quality of life.(5) They discounted with 4.5% per year.(1)

Cost-effectiveness is not the only argument to implement or continue a screening programme. There is a need for public support based on proper information on harms and benefits. We estimated breast cancer mortality reduction of up to 30% and an acceptable overdiagnosis rate.(12)

Crude breast cancer mortality rate of women of all ages had a sharp decline in 1994, which cannot be satisfactorily explained by mammography use. The drop in mortality reduction is too soon after the introduction of mammography outside the NBCSP and too large, given the fact that also opportunistic screening has gradually increased, to be the result of opportunistic screening. This decline is probably the result of adjuvant therapy and improvements in overall survival rates, possibly enhanced by the “re-organization of the

breast cancer health care system".(37) In the second half of the 1980s and the first half of the 1990s combined multi-agent chemotherapy and Tamoxifen were being used more frequently.(38) Treatment effects are included in our model. The fact that breast cancer mortality keeps decreasing despite increases in incidence, indicates that over the years following the sharp decline multiple factors have contributed to this decrease.

We found an estimated maximal mortality reduction in 2022 of 30% increasing breast cancer mortality rate for the Norwegian population, gradually having been exposed to screening. This is in line with the estimated breast cancer mortality reduction of approximately 28% found by Weedon-Fekjaer.(39) In the model we used the estimated effects of screening and treatment to establish mortality reduction. In reality mortality reduction is even greater. Because we do not know what causes the greater reduction in mortality we used a conservative estimate of mortality reduction based on screen effects and treatment effects.

The analysis is based on the outcomes of the MISCAN model. There are some limitations to the model. First the model assumes an increased background risk for breast cancer, which cannot be fully explained.(2, 3, 12) Second we performed analysis on a cohort run. The benefits of a cohort run is that all women in the model are the same age, and that follow up for the entire population is complete. The drawback of a cohort run is that it may overestimate the effects of a program, because in an actual population not all women are exposed to screening at the same age, and follow up is never complete for all women at the same time.

Another limitation of the study is that the data on utilities is from a period when treatment was different from now. Recent figures are not readily available.

In conclusion, we estimate a breast cancer mortality reduction of 16% in women aged 55-80 years, with a projected maximal reduction in 2022 of 30%. The NBCSP is cost-effective in preventing breast cancer specific mortality.

REFERENCES

1. Hofvind S, Wang H, Thoresen S. Do the results of the process indicators in the Norwegian Breast Cancer Screening Program predict future mortality reduction from breast cancer? Acta Oncol. 2004;43(5):467-73.
2. Duffy SW, Michalopoulos D, Sebuodegard S, Hofvind S. Trends in aggregate cancer incidence rates in relation to screening and possible overdiagnosis: A word of caution. J Med Screen. 2014;21(1):24-9.
3. 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.
4. Norum J. Breast cancer screening by mammography in Norway. Is it cost-effective? Ann Oncol. 1999;10(2):197-203.
5. Wang H, Karesen R, Hervik A, Thoresen SO. Mammography screening in Norway: results from the first screening round in four counties and cost-effectiveness of a modeled nationwide screening. Cancer Causes Control. 2001;12(1):39-45.
6. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2013;6:CD001877.
7. Kalager M, Zelen M, Langmark F, Adami HO. Effect of screening mammography on breast-cancer mortality in Norway. N Engl J Med. 2010;363(13):1203-10.
8. 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(7):491-9.
9. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56.
10. 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(1):111-21.
11. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen. 2012;19 Suppl 1:33-41.
12. Research-based evaluation of the Norwegian Breast Cancer Screening Program. The Research Council of Norway, May, 2015. Report No.
13. IARC handbooks of cancer prevention, volume 15. Breast cancer screening.: International agency for research on cancer; 2015.
14. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol. 2010;49(5):725-36.
15. Jones RH, Dey I. Determining one or more change points. Chem Phys Lipids. 1995;76(1):1-6.
16. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985;20(1):79-93.
17. Habbema JD. A simulation approach to cost-effectiveness and cost-benefit calculations of screening for the early detection of disease. Eur j oper res. 1987;29(2):159-66.

18. de Gelder R, Bulliard JL, de Wolf C, Fracheboud J, Draisma G, Schopper D, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. Eur J Cancer. 2009;45(1):127-38.
19. Sukel MP, van de Poll-Franse LV, Nieuwenhuijzen GA, Vreugdenhil G, Herings RM, Coebergh JW, et al. Substantial increase in the use of adjuvant systemic treatment for early stage breast cancer reflects changes in guidelines in the period 1990-2006 in the southeastern Netherlands. Eur J Cancer. 2008;44(13):1846-54.
20. de Gelder R, Heijnsdijk EA, Fracheboud J, Draisma G, de Koning HJ. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. Int J Cancer. 2015;137(1):165-72.
21. de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst. 1995;87(16):1217-23.
22. Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am. 2000;38(4):625-51.
23. Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, et al. Mammography activity in Norway 1983 to 2008. Acta Oncol. 2011;50(7):1062-7.
24. FolkeHelseInstituttet. Norwegian Prescription Database http://wwwnorpdno/. 2004;2013.
25. van Luijt PA, Heijnsdijk, E.A.M., van Ravesteyn, N.T., Hofvind, S., de Koning, H.J. Breast cancer incidence trends in Norway and estimates of overdiagnosis. J Med Screen. 2016;in press.
26. Moger TA. Direct and indirect costs of the Norwegian Breast Cancer Screening Program. . HERO On line Working Paper Series

[Internet]. 2012; 3. Available from: http://www.med.uio.no/helsam/forskning/nettverk/hero/publikasjoner/skriftserie/2012/hero2012-3.pdf.
27. Hofvind S, Geller B, Vacek PM, Thoresen S, Skaane P. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. Eur J Epidemiol. 2007;22(7):447-55.
28. Seo BK, Pisano ED, Kuzmiak CM, Koomen M, Pavic D, McLelland R, et al. The positive predictive value for diagnosis of breast cancer full-field digital mammography versus film-screen mammography in the diagnostic mammographic population. Acad Radiol. 2006;13(10):1229-35.
29. Katalinic A, Bartel C, Raspe H, Schreer I. Beyond mammography screening: quality assurance in breast cancer diagnosis (The QuaMaDi Project). Br J Cancer. 2007;96(1):157-61.
30. Moger TA, Bjornelv GM, Aas E. Expected 10-year treatment cost of breast cancer detected within and outside a public screening program in Norway. The European journal of health economics : HEPAC : health economics in prevention and care. 2016;17(6):745-54.
31. de Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van Der Maas PJ. The impact of a breast cancer screening programme on quality-adjusted life-years. Int J Cancer. 1991;49(4):538-44.
32. NICE. NICE guide to the methods of health technology appraisal. London: 2008.
33. Cost-effectiveness and strategic planning (WHO-CHOICE) [10/08/2014]. Available from: http://www.who.int/choice/costs/CER_thresholds/en/.

34. Bank W. GDP 2014 [Available from: http://data.worldbank.org/indicator/NY.GDP.PCAP.CD.
35. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness--the curious resilience of the $50,000-per-QALY threshold. N Engl J Med. 2014;371(9):796-7.
36. 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(11):2205-40.
37. Kalager M, Haldorsen T, Bretthauer M, Hoff G, Thoresen SO, Adami HO. Improved breast cancer survival following introduction of an organized mammography screening program among both screened and unscreened women: a population-based cohort study. Breast Cancer Res. 2009;11(4):R44.
38. Harlan LC, Abrams J, Warren JL, Clegg L, Stevens J, Ballard-Barbash R. Adjuvant therapy for breast cancer: practice patterns of community physicians. J Clin Oncol. 2002;20(7):1809-17.
39. Weedon-Fekjaer H, Romundstad PR, Vatten LJ. Modern mammography screening and breast cancer mortality: population study. BMJ. 2014;348:g3701.

CHAPTER 7:

GENERAL DISCUSSION

RESEARCH QUESTIONS AND OUTLINE OF THIS THESIS:

THE DETECTION OF DCIS SEEMS TO BE ASSOCIATED WITH A HIGHER RISK OF OVERDIAGNOSIS. WHAT IS THE IMPACT OF THE TRANSITION TO DIGITAL MAMMOGRAPHY ON THE AMOUNT OF DCIS DIAGNOSED WITH SCREENING (CHAPTER 2)?

We found that during and after the transition of the Dutch breast cancer screening program from film screen mammography to digital mammography, the program had a higher detection level. Significantly more DCIS and very small tumors were found with digital mammography.

An earlier modeling study showed that digital mammography screening would further reduce breast cancer mortality by 4.4%, at a 21% increased overdiagnosis rate.(1)

To put these results in an international perspective: In 2015 the International Breast Cancer Screening Network (IBCSN) published a paper reporting performance indicators of 15 different regions, data were collected between 2004 and 2008.(2) The Netherlands has a low detection rate compared to Norway and Denmark, but at the same time an average detection rate when compared to the overall rate of the 15 regions that had provided data to the report. The proportion of DCIS in the Dutch program is significantly higher than the average of the regions in the report, but lower than the percentage of DCIS in Norway (Table 1). Recall rate was not mentioned in the report.

We also compared detection rate, recall rate and percentage of DCIS in the Dutch program to data from the United Kingdom, as provided in the NHS evaluation report.(3) In the United Kingdom 90% of all screening units had a full field digital mammography set in 2009. The detection rate in the UK program in 2009, for women aged 45-74, was also higher than it was in the Netherlands. Recall rate was 8.2/1,000 screens in 2009 in women aged 45-74, compared to an average of 5.9/1,000 screens in women aged 50-74 in the Netherlands between 2004 and 2010 with digital mammography.

The detection rate is not only dependent on the sensitivity of the screening program, but on the background incidence of breast cancer as well. In countries with a higher incidence of breast cancer, the program will detect more breast cancer.

The Dutch program managed to achieve an average detection rate even with relatively low recall rates. The Dutch program detects an average amount of DCIS.

Table 1. Detection rate (DR) per 1,000 screens and proportion of DCIS in percentages in the Netherlands, compared to results in the Danish, Norwegian and UK program. IBCSN: International Breast Cancer Screening Network, NHS: National Health Services.

	DR	95% CI	% DCIS	95% CI
IBCSN				
Denmark	7.13	(6.70 ; 7.57)	13.20%	(11.14% ; 15.27%)
Norway	5.27	(5.13 ; 5.42)	17.82%	(16.76% ; 18.87%)
Overall (IBCSN 2015)	4.90	(4.85 ; 4.95)	15.24%	(14.87% ; 15.62%)
NHS evaluation report				
United Kingdom	8.17	(8.05 ; 8.30)	20.00%	(19.39% ; 20.61%)
The Netherlands	4.90	(4.73 ; 5.06)	16.39%	(15.16% ; 17.61%)

OVERDIAGNOSIS IS A WELL DEBATED ISSUE IN BREAST CANCER SCREENING, BUT NOT IN SCREENING FOR CERVICAL CANCER SCREENING. IS THERE OVERDIAGNOSIS IN CERVICAL CANCER SCREENING, AND HOW DOES THIS RELATE TO THE OVERDIAGNOSIS RATE IN BREAST CANCER SCREENING (CHAPTER 3)?

We calculated estimates for overdiagnosis rate in cervical cancer screening and in breast cancer screening and found that the cervical cancer screening also yields quite some overdiagnosis. In breast cancer screening the detection of pre-invasive disease (DCIS) adds to the incidence of breast cancer. In cervical cancer screening the detection of pre-invasive disease (CIN) is not counted in the incidence of cervical cancer. Both DCIS and CIN warrant treatment. The treatment of pre-invasive disease aims to prevent the occurrence of advanced disease and cancer related mortality.

In breast cancer screening DCIS constitutes an early phase of breast cancer. Not all DCIS will progress to invasive disease within a woman's lifetime. The mortality rate from breast cancer in women with DCIS is low.(4) Treatment of DCIS however is extensive and includes (partial) mastectomy and sometimes radiation therapy. The burden of overdiagnosis in breast cancer is considered substantial.

In the cervical cancer screening program CIN is practically the only disease detected with screening.(5) In cervical cancer, pre-invasive disease is not considered a cancer diagnosis. The detection and elimination of pre-invasive disease prevents the occurrence of cancer at a later stage in life. One could argue that this is a late form of primary prevention. Even in the setting of cervical cancer screening however, some women will be diagnosed and treated unnecessarily. Some of the CIN lesions detected at population-based screening would have never developed into cervical cancer, and never become clinically apparent in the woman's

lifetime. The treatment for CIN is not very invasive and the occurrence of complications is rare. However, treatment for CIN may be associated with preterm delivery, low birth weight, caesarean section and preterm rupture of the membranes in future pregnancies.(6) Because the number of lesions detected and treated is substantial, 5,037 screen detected CIN/ 100,000 women, this still constitutes a substantial burden.

WHAT IS THE IMPACT OF DCIS ON OVERDIAGNOSIS ESTIMATES IN BREAST CANCER SCREENING (CHAPTER 4)?

We found that the distribution of DCIS differentiation grades does not differ between screened women and women who had never attended population-based screening. The majority of DCIS (50.9%) is high grade. The overdiagnosis rate of DCIS by grade on a population level varies from 60% in low-grade DCIS, to 56% in intermediate-grade DCIS, and 45% in high-grade DCIS. The overdiagnosis rate of DCIS by grade on an individual level varies from 61% in low-grade DCIS, to 57% in intermediate-grade DCIS, and 45% in high-grade DCIS.

The impact of detecting more DCIS with new screening strategies; i.e. digital mammography, and in the future possibly tomosynthesis, on overdiagnosis is significant. The best way to address this expected increase would be to tailor the treatment of DCIS to expected natural behavior. New strategies to predict natural behavior will include genetic profiling.

IS THE NORWEGIAN BREAST CANCER SCREENING PROGRAM GENERATING OVERDIAGNOSIS, AND AT WHAT LEVEL (CHAPTER 5)?

We found that breast cancer incidence in Norway has been erratic. Many factors have been identified that influence breast cancer incidence, mostly hormone replacement therapy (HRT) and the increased use of mammography screening. These factors together, mammography use and HRT use, cannot satisfactorily explain the erratic course of breast cancer incidence in Norway. We were able to replicate the increase by increasing the background incidence with an increased relative risk, applicable to all women in the population under age 87, of 1.75, in the model. The added relative risk correlates nicely to the work of Weedon-Fekjaer et al. who found a similar relative risk in the cohort effect, increasing the RR from 0.74 to 1.25 (a factor of 1.7) to fit his age-period-cohort model to the observed data.(7) The difficulties in explaining the erratic course of breast cancer incidence have also already been pointed out by Duffy et al.(8)

The cause of this large increase in background incidence remains unknown. Some argue that all of this increase must be the result of overdiagnosis in the NBCSP.(9) Our model has always been able to predict incidence rates of breast cancer very well in the Netherlands and also in other countries (Germany, Spain, US).(10-12) Therefore we have confidence in our disease model. We assume that the disease process from onset is the same in Norwegian women with breast cancer as in Dutch women with breast cancer. The earliest steep increase in breast cancer incidence occurred in 1995, prior to the introduction of screening. Therefore it is more plausible that the rate of onset is higher, than that the increase is incidence would be attributable to overdiagnosis.

We estimate an overdiagnosis rate of 2% from a population perspective (number of excess diagnoses in a situation with screening divided by the total number of diagnoses in a situation with screening (screendetected and clinically detected) in women aged 50 to 100), and 3% from an individual perspective (number of excess diagnoses in a situation with screening divided by the total number of diagnoses in a situation with screening (screendetected and clinically detected) in women aged 50 to 75). This corresponds with overdiagnosis estimates in different countries, and is well within the range of what is deemed acceptable internationally.(13)

The breast cancer mortality reduction caused by the NBCSP will increase in coming years, up to 105 breast cancer deaths prevented annually among women aged 55-80 in 2030. In accordance to the findings of Marmot et al. we found that a follow-up period of at least 10 - 15 years is necessary to appreciate the full impact of the screening on breast cancer mortality (Figure 1).(13)

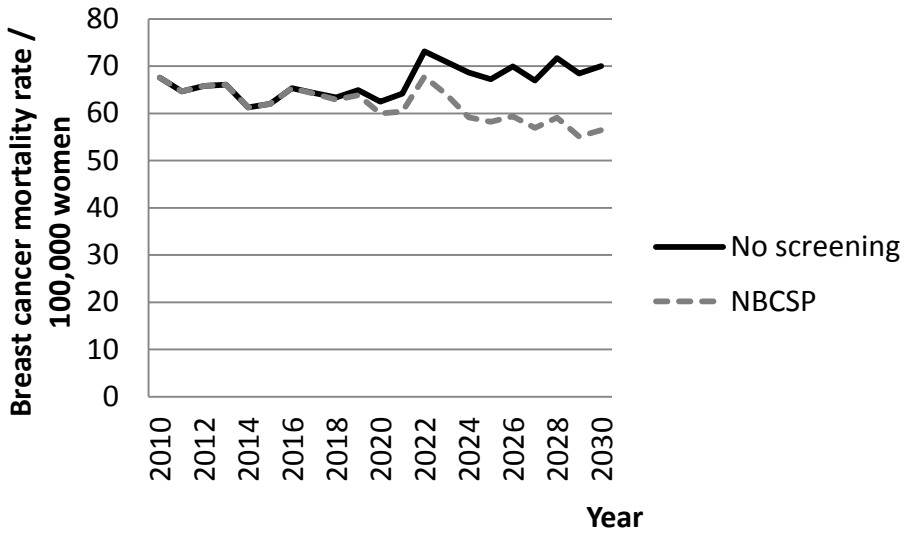


Figure 1. Estimated breast cancer mortality reductions in Norway in 2010 to 2030 per 100,000 women aged 55-80 invited to screening.

IS THE NBCSP EFFECTIVE IN REDUCING BREAST CANCER MORTALITY DESPITE THE OCCURRENCE OF OVERDIAGNOSIS (CHAPTER 6)?

We found that the NBCSP is highly cost-effective. The program achieved a breast cancer mortality reduction of 16% in a cohort of women aged 49 in 2004, with complete follow-up. The number of screen examinations necessary to prevent one breast cancer death (number needed to screen, NNS) is 1,470. The incremental costs per QALY gained is 112,162 NOK (11,884 EUR) for direct medical costs of screening, or 189,557 NOK (20,084 EUR) for direct medical and indirect costs of the screening.

If we would assume that the increased breast cancer incidence in Norway is the result of overdiagnosis, the impact of a screening program would be less favorable. We can model this situation by assuming a high prevalence of slow growing tumors, or of tumors with a high potential for regression. If there is a large pool of dormant disease with little clinical significance, the detection of this disease will yield more overdiagnosis, without reducing breast cancer mortality.

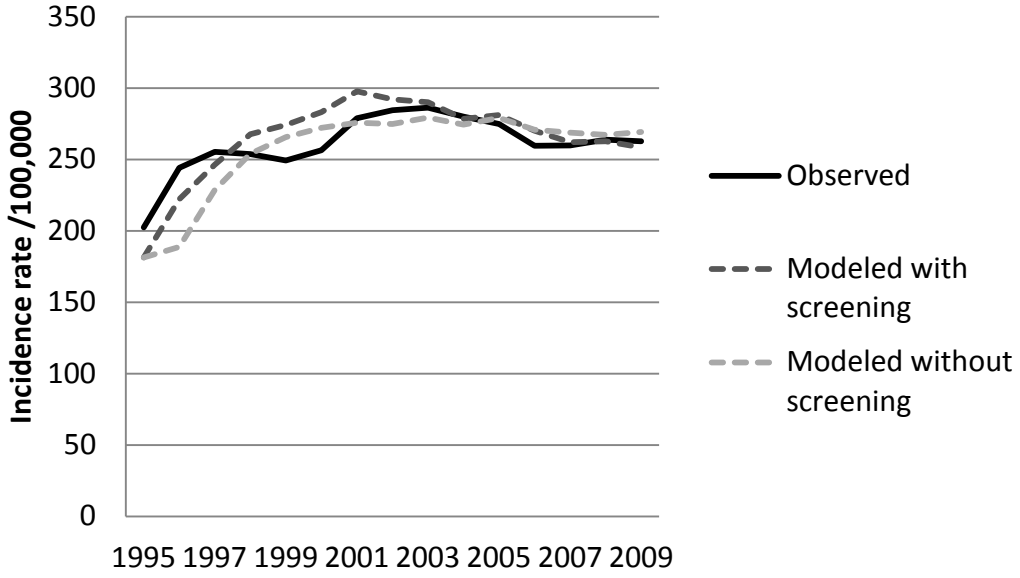
We modeled this situation by increasing the time of progression between onset of DCIS to early pre-clinical invasive disease (Table 3). Because the possibility of progression to the next stage is dependent on the probability of progression of the previous state, this will increase the time to progression between subsequent disease states as well. Now the model will predict a large amount of overdiagnosis and less prevented deaths per screen. We can use this model to calculate the most conservative estimate of the cost-effectiveness of the NBCSP (Table 4).

Table 3. Average duration of the stages (in years) for women aged 50 years in the model with the added relative risk and in the model with the exceptionally long dwell times.

	High relative risk	Long dwell time
DCIS	2.09	25.00
T1a	0.09	0.01
T1b	0.37	0.07
T1c	0.82	0.76
T2+	0.89	0.65

This scenario does not fit the observed incidence data as well as our added relative risk model, and is only provided for the purpose of a conservative estimate (Figure 2). Even in this scenario, the program is cost-effective considering the WHO guidelines (threshold 1,926,366 NOK).

A



B

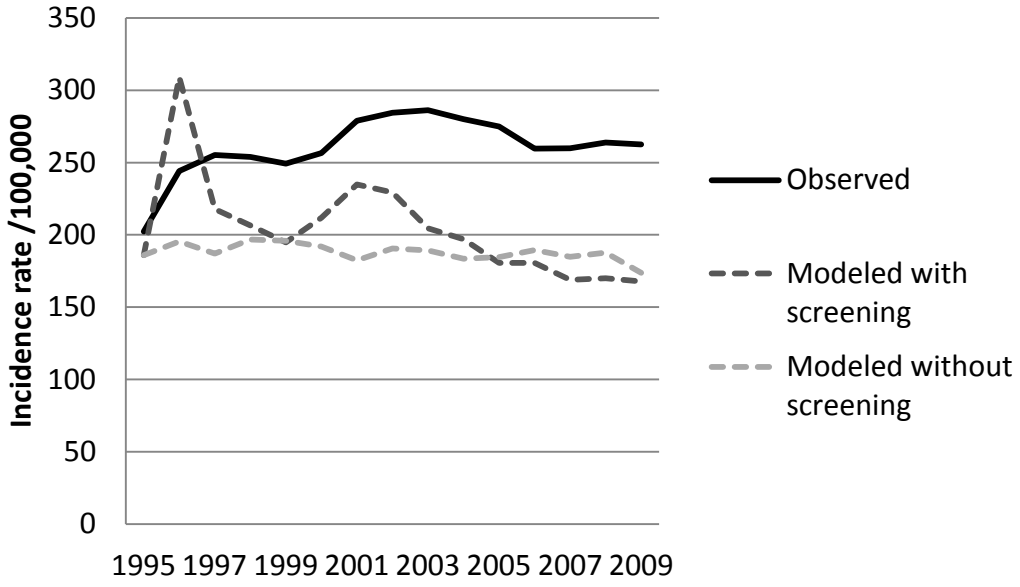


Figure 2: Breast cancer incidence rate/ 100,000 women aged 50-75 in Norway. A) Model with added relative risk, B) Model with extended dwell times to mimic a maximal amount of overdiagnosis.

Table 4. Effects, costs and breast cancer mortality reduction per 100,000 women, aged 49 in 2004, with complete follow-up (to 2055), with screening only in the NBCSP, in a scenario where we ascribed most of the increase in breast cancer incidence to overdiagnosis. QALY: quality adjusted life years. Only the incremental costs/QALY gained are given with a 3.5% annual discount.

	Without screening	With screening
Screening tests	-	439,142
Health effects		
Cancer diagnosed	4,249	4,933
Screen detected cancers	-	2,018
False positives	7,224	21,097
Breast cancer deaths	785	641
QALY gained	-	1,120
Mortality reduction	-	18%
Number needed to screen	-	3,042
Costs (in NOK x1000)		
Direct medical costs of screening	-	356,583
Total costs of screening	-	554,197
Diagnosis	32,916	74,681
Treatment	1,046,853	1,015,458
Total costs based on direct medical costs of screening	1,079,769	1,446,722
Total costs based on total costs of screening		1,644,336
Incremental costs based on direct medical costs of screening	-	366,953,096
Incremental costs based on total costs of screening		564,566,996
Cost-effectiveness		
Direct medical costs of screening		
Incremental costs/QALY gained, discounted in NOK		327,637
Total costs of screening		
Incremental costs/QALY gained, discounted in NOK		504,078

GENERAL DISCUSSION

METHODOLOGY

Many objections have been raised to modelling studies, arguing that the assumptions do not hold, or are not communicated transparently enough.(14)

Modelling studies can compare hypothetical situations. This comparison is necessary in the evaluation of breast cancer screening, because screening is so widely adopted. There are no identical populations available that are not exposed to population-based screening. Comparisons have been made between historical groups, comparing the situation prior to the introduction of breast cancer screening to the current situation with breast cancer screening.(15) The disadvantage of this approach is the introduction of bias, caused by the fact that other circumstances have also changed in time (i.e. incidence, treatment, awareness, overall survival). Another way to address the problem is by comparing different countries that have introduced population-based screening at a different time.(16) In this comparison bias can occur because different countries have different overall health policies and overall survival rates. Also, in countries surrounded by countries that have a population-based screening program in place, women are bound to have abundant opportunistic screening.

The MISCAN model was first developed in 1985.(17) Since then the model has repeatedly been recalibrated with the latest available data, provided by the Netherlands Evaluation of the Breast cancer screening program Team (NETB) and the Dutch cancer registry. These calibration processes have always been reported and documented.(1, 18) Predictions made by the model in 1995 on breast cancer incidence rates in 2010 closely resemble the observed breast cancer incidence rate observed in 2010.(19) The assumptions made in modeling the disease process and the impact of screening have been extensively documented.(1) This documentation has been so precise that independent researchers have been able to closely replicate the results by recreating the model from the descriptions.(20) The details of the current MISCAN model can be found in appendix 1.

The evidence derived from modeling studies can never compete with data provided by randomized controlled trials. The data provided by such trials however are incorporated in the model. Modeling studies provide us with a validated, reliable insight on the current performance of the breast cancer screening program, allow us to continually monitor if the screening program is still effective, and can translate screening trial results to different settings or screening protocols. Another advantage of a model is the ability to model complete follow-up.

DIGITAL MAMMOGRAPHY UPDATE

Some time has passed since the evaluation of the transition from film screen to digital screen mammography in chapter 2. We compared our results from the years 2004-2010 to the now complete data from 2013.

In the period 2010-2013, in women who were screened for the first time (aged 49-74), recall rate increased to 6.3%. This resulted in a high detection rate in these women of 8.5/1,000 women screened. But the positive predictive value (PPV) remained low at 13.5% (Table 4). In women who were screened at a subsequent visit, within 2.5 years since their last visit (timely subsequent screen), recall rate remains low at 2.0%, with a detection rate of 8.5/1,000 women screened, resulting in a PPV of 32.3% (Table 5 and 6).

The proportion of DCIS in all detected malignancies is high in first screens with 31%, which is higher than it was in 2004-2010 (25%). Among malignancies detected at timely subsequent screens the proportion of DCIS is unchanged at 21%.

Table 5. Recall rate, detection rate, and positive predictive value of the Dutch screening program in 2013 and in 2004-2010 in women screened for the first time. PPV: positive predictive value.

	2004-2010		2013	
	Point estimate	95% CI	Point estimate	95% CI
Recall rate (%)	2.8	(2.70 ; 2.80)	6.3	(6.28 ; 6.37)
Detection Rate (per 1,000 women screened)	5.2	(4.99 ; 5.41)	8.5	(8.39 ; 8.70)
PPV (%)	20.8	(20.03 ; 21.52)	13.5	(13.45 ; 13.57)

Table 6. Recall rate, detection rate, and positive predictive value of the Dutch screening program in 2013 and in 2004-2010 in women with a timely subsequent screening (within 2.5 years since the previous screening). PPV: positive predictive value.

	2004-2010		2013	
	Point estimate	95% CI	Point estimate	95% CI
Recall rate (%)	1.7	(1.72 ; 1.77)	2.0	(1.97 ; 2.02)
Detection Rate (per 1,000 women screened)	6.2	(6.07 ; 6.37)	6.5	(6.31 ; 6.58)
PPV (%)	32.5	(34.73 ; 36.14)	32.3	(32.20 ; 32.36)

The recall rates and detection rates vary among different groups of reading radiologists. An analysis of the period 2008-2011 showed that even though there were reading groups that differed significantly from the national average for every parameter (recall rate, detection rate and PPV), both positively and negatively, there was only one reading group of the 16 groups that performed significantly worse on all parameters than the national average.(21)

Recall decisions are based on the score in the Breast Imaging Reporting and Data System (BI-RADS) classification system. Every examination is read by two different radiologists. If there is a difference in BI-RADS code, the two readers can confer with one and other to reach consensus, or call in a third party for arbitration. The radiologists that reports the latest BI-RADS code, decides.

There are seven different BI-RADS codes: 0: there is a lesion, further investigation necessary to determine probability of malignancy; 1: normal; 2: benign finding; 3: probably benign finding, re-evaluate after 6 months, 4: lesion, suspected malignancy, not typical for malignancy, 5: probable malignancy, 6: histologically proven malignancy. For obvious reasons BI-RADS 3 and 6 are not applicable in the screening situation.

The outcome of screening in terms of BI-RADS were; 52.7% of all recalls are the result of a BI-RADS 0. The PPV of a recall based on a BI-RADS 0 is 11.5%. The amount of recalls based on BI-RADS 0 increased in the years 2010-2013 from 43.9% in 2010 to 56.4% in 2013. On the other hand, the amount of recalls based on BI-RADS 4 decreased in this period from 49.2% to 37.5%. The percentage of DCIS in women recalled for a BI-RADS 0 decreased from 11.5% in 2010 to 8.2% in 2013, when small invasive carcinoma with negative lymph nodes in these women increased from 61.1% to 65.8%.(22)

So radiologists have found more lesions that warrant further investigation, rather than lesions that are probably malignant, but at the same time these lesions were more often invasive carcinomas.

BREAST CANCER INCIDENCE IN NORWAY FROM AN INTERNATIONAL PERSPECTIVE

Because the situation in Norway appears exceptional in terms of breast cancer incidence increase, we compared the observed data from Norway in the years 1990-2013 to Dutch data from 1990-2013, Finnish data from 1989 -2013, and Swedish data from 1989-2013.

Invasive breast cancer incidence in 2003 did not differ much, the most remarkable was the increase of breast cancer incidence. In the Netherlands invasive breast cancer incidence in women aged 50-70 years old rose from 228 /100,000 women in 1990 to 309 /100,000 women in 2003.(23) In Norway invasive breast cancer incidence in 1990 was only 158 /100,00 women, yet rapidly increased up to 298 /100,000 women in 2003 (Figure 3).(24) The increase in invasive breast cancer incidence was therefore twice as high. After 2003, there was some decline in invasive breast cancer incidence in women aged 50-70 years in Norway. At the same time the invasive breast cancer incidence in the Netherlands kept increasing. This may be explained by the reduction in the number of breast cancers caused by the reduced use of HRT in Norway.

Invasive breast cancer incidence in Finland rose from 213 /100,000 women in 1989 to 297 /100,000 women in 2003, but kept increasing up to 336 /100,000 women in 2013.(24)

Invasive breast cancer incidence in Sweden started at the same level in as the Netherlands in 1989 with 236 invasive breast cancers/ 100,000 women, reached 305 /100,000 women in 2003 and remained steadily around 300 /100,000 women up to 2013 (Figure 3). Therefore, the increase in Norway is still by far the highest.(24)

One could argue that breast cancer registration in the early years was incomplete. This would explain a rather low breast cancer incidence in the years prior to 1990 and the rapid catch up in the years after that. It cannot explain the decrease in breast cancer incidence since 2003 however. Moreover the Norwegian Cancer Registry has a reported 100% coverage of incident cases for the entire period.

We know that the use of HRT in Norway has been exceptionally high in the 1990's.(25, 26) After publication of the Million Women study however, HRT use rapidly declined. Although some decline in breast cancer incidence can be seen from 2006, it does not return to the situation prior to the massive HRT intake. The incidence rate after 2003 is approximately 270 /100,000, in comparison to 160 /100,000 in the early 1990's.

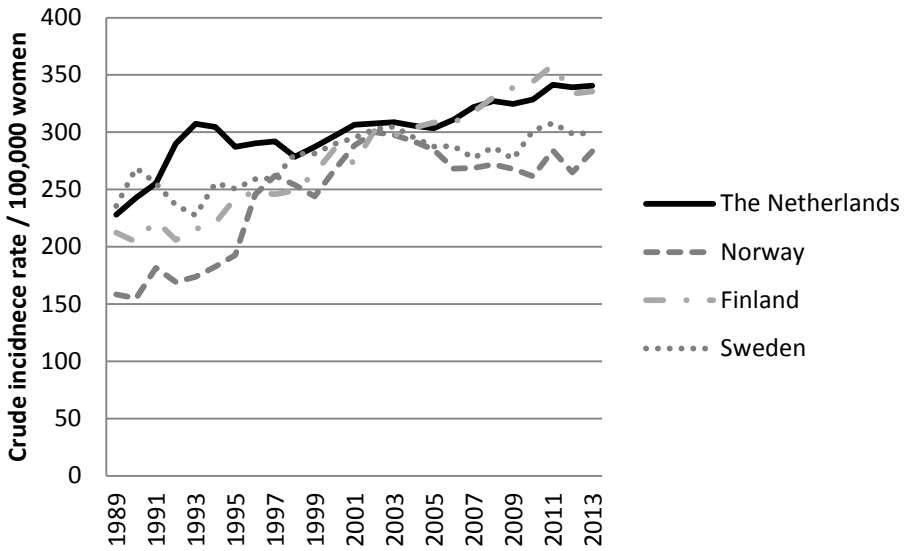


Figure 3. Invasive breast cancer incidence per 100,000 women aged 50-70 in The Netherlands, Norway, Finland and Sweden from 1989 to 2013.

BREAST CANCER MORTALITY AFTER THE INTRODUCTION OF SCREENING

In the Netherlands breast cancer mortality in women aged 55-80 has been decreasing steadily since 1990. This coincides with the start of population-based mammography screening. However, we cannot attribute this reduction to screening alone, given the fact that we expect mortality reduction to take effect after at least five years of follow-up. In Norway a similar pattern can be seen; breast cancer mortality reduced gradually after 1997, which is too early to be fully attributable to screening alone (Figure 4). The decrease in breast cancer mortality in both countries is probably the combined effect of early detection and advances in adjuvant treatment regimens.(18)

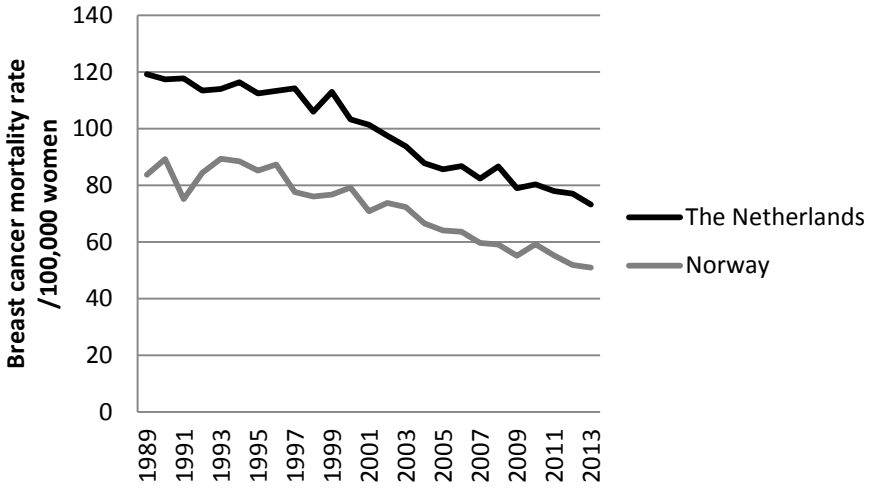


Figure 4. Breast cancer mortality rate per 100,000 women aged 55-80 in the Netherlands and Norway.

DCIS, REDUCING THE IMPACT OF OVERDIAGNOSIS

More and more diagnoses are DCIS. The mortality rate from breast cancer after DCIS is very low.(4) A recent analysis of the 10- and 20-year mortality from DCIS based on 108,196 women with DCIS from the Surveillance, Epidemiology, and End Results (SEER) database from the United States, showed that the 10-year breast cancer specific mortality rate after a diagnosis of DCIS was 1.1%. The 20-year breast cancer specific mortality rate after a diagnosis of DCIS was 3.3%.

A large proportion of DCIS represents overdiagnosis. This is a problem, both from an individual point of view; women will suffer unnecessary treatment and anxiety, and from a population point of view; resources allocated to medical care are being wasted.

The solution to this problem is not to stop looking for DCIS. The solution should be to be able to discriminate between those DCIS that will progress to aggressive invasive disease and those DCIS that will remain dormant, or even regress back to normal. More and more diagnostic tools are becoming available in an attempt to predict natural behavior. In the near future a new study will be conducted to include biomarkers in the determination of low-risk and high-risk DCIS.

Individualized policies towards treatment need to be developed based on the expected natural behavior of each DCIS. In order to develop such strategies, two large trials have been set up in recent years.

The LORIS trial, funded by the National Health Services (NHS), in the United Kingdom aims to randomize women aged 46 and older, with screen-detected or incidental calcifications on mammography, that constitute a histologically proven non-high grade DCIS on core biopsy, vacuum assisted core biopsy, or surgical biopsy, into a follow-up arm, with annual mammography for 10 years, and a surgery arm, followed by annual mammography for 10 years.(27) Inclusion appears to be difficult because patients do not like to be randomized for surgery or no surgery.(28)

The second trial is the LOw Risk DCIS (LORD) trial, a European trial, funded by the European Organization for Research and Treatment of Cancer (EORTC). It aims to randomize “women aged 45 and older with asymptomatic, pure and low-grade DCIS based on vacuum assisted biopsies of calcifications only, detected by population-based population-based screening or opportunistic screening mammography” into standard care (usually surgery), followed by annual mammography for 10 years, and active surveillance, which consists of annual mammography for 10 years.(29)

Other initiatives include a trial by Hwang et al., to be set up in the near future in the United States, in response to recent findings in the trend in treatment of DCIS in the US.(30)

One of the reasons why it is so very hard to include women in a watchful waiting trial for low-grade DCIS could be that women prefer surgery because DCIS is considered an early form of cancer. It seems contradictory to diagnose someone with cancer, and not treat them. Some experts have already argued that it might be beneficiary to change the name of DCIS, to avoid the word “cancer”.

FUTURE DIRECTIONS

TOMOSYNTHESIS

New imaging strategies in breast cancer imaging include tomosynthesis. Tomosynthesis uses a machine comparable to mammography to obtain not just one image of the breast, but several images, each at a slightly different angle between X-ray source and detector. This produces a stack of images, through which the radiographer can scroll. The technique allows to discriminate between real masses and over-projection. The images can be combined to produce a C-view image, on which calcifications are better appreciated.

As of this moment tomosynthesis is only used in a clinical setting, but because the machine uses approximately the same amount of radiation as a normal mammogram does, it is being considered to replace traditional mammography in a screening setting. Two trials have been conducted to assess the feasibility of tomosynthesis in a screening situation. They found that the use of tomosynthesis in combination with 2D imaging improves breast cancer detection and has the potential to reduce false positive recalls.(31, 32) The costs of storing the extra amount of images associated with tomosynthesis is dependent on the degree of image compression allowed, these costs would be added to the €65,- per examination in the current situation (preliminary report of the NETB 2016).

Tomosynthesis is estimated to have a higher sensitivity for both masses and calcifications. Therefore implementation of tomosynthesis in a screening setting would increase early diagnosis at the cost of more overdiagnosis. How much breast cancer mortality reduction can be obtained and at the cost of what amount of overdiagnosis needs to be calculated.

RISK STRATIFICATION

In search of an individualized approach to breast cancer screening, two trials are being conducted to stratify women with a different risk of having breast cancer in different screening strategies.

The first trial is the DENSE trial, in which women in the breast cancer screening program with extremely dense breasts (ACR 4) and a negative mammogram are randomized in either the standard care arm (no additional mammography), or in the arm in which they undergo additional MR examination with dynamic MRI with gadolinium based contrast medium for three screening rounds. The aim is to evaluate the value of additional MRI in women with dense breasts. Dense breasts are associated with a lower sensitivity of mammography, and a higher risk of developing breast cancer.(33)

The second trial is the PRISMA trial, where participants of the breast cancer screening program are asked to fill out an online questionnaire, for permission to measure breast density on the mammogram, and for a blood sample. The aim is to determine risk factors,

that increase the probability of developing breast cancer and to design screening strategies based on these risk factors.(34)

CO-MORBIDITY

Women with a high co-morbidity may benefit from a less stringent screening strategy, as their mortality is determined by their co-morbidity rather than a small screen-detected breast cancer. Women who have very little or no co-morbidity may benefit from breast cancer screening despite advanced age. In order to be able to tailor screening strategies to co-morbidity, women with high co-morbidity need to be identified and strategies need to be evaluated, for example by using the MISCAN model.

CONCLUSIONS

The transition to digital mammography has increased breast cancer detection level, mostly of DCIS and small invasive disease.

The cervical cancer screening program yields even more overdiagnosis than the breast cancer screening program, but less invasive treatment is warranted.

DCIS grade distribution does not differ between screened women and women who never attended population-based screening. Higher grade DCIS has a lower overdiagnosis estimate. The contribution of DCIS to the overdiagnosis estimate is high.

Breast cancer incidence in Norway is erratic, and cannot be fully explained by the use of HRT and mammography alone. Overdiagnosis estimates of the NBCSP are low, and well within international limits.

Breast cancer screening is a safe and cost-effective way to prevent breast cancer deaths in Norway.

Overdiagnosis in breast cancer screening occurs, but the occurrence does not outweigh the benefits of population-based screening.

RECOMMENDATIONS FOR FURTHER RESEARCH

The changes in breast cancer in Norway, which cannot be fully explained by current knowledge about risk factors, warrant further investigation. It would be interesting to know some more characteristics of the tumors, and if these characteristics have changed over time, along with the incidence changes.

Further and more specific risk stratification of DCIS and small invasive tumors is needed to adequately predict natural behavior and necessary treatment. Better individual management of DCIS and small invasive tumors can lower the impact of overtreatment and further research should be aimed at the least invasive treatment strategies with favorable outcome for these patients.

To decrease overtreatment and reduce the amount of invasive diagnostic tests, we need to be able to determine malignancy and grade on imaging alone, without the need of invasive biopsies for BI-RADS IV lesions. Current practice for calcifications is to do a vacuum assisted biopsy, often leading to a diagnosis of DCIS. If it were possible to definitely determine low-grade DCIS on imaging alone, and if it would have been proven safe to manage this with a watchful waiting policy, then an overdiagnosed low-grade DCIS would be of much less impact. To obtain this amount of certainty about diagnosis and grade on imaging alone, research would have to be done on tomosynthesis imaging, ultrasound imaging and MRI, to determine imaging characteristics only seen in low-grade DCIS.

With increased life expectancy the upper age limit of screening might have to be re-evaluated. Any endeavor in this direction needs to be accompanied with a more personalized approach. Not all women over 75 are the same. Some are in the prime of their life, some have extensive co-morbidity. For those expecting to live to be over 100 years old, screening may add a significant amount of life-years. The implications of expanding the program need to be calculated, so that cost-effectiveness can be assessed.

Already much work has been done on the implications of lowering the age-limit of population-based screening.(35) Despite arguments that this could be cost-effective, it has not been implemented yet. A pilot program to objectively observe the effects of lowering the age-limit is necessary to gain insight in the actual implications.

Population-based screening for breast cancer is here to stay, and for good reasons. The challenge will be to adequately address the occurrence and consequences of the generated overdiagnosis by tailoring treatment and follow-up regimes to the personal needs of patients.

REFERENCES

1. de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, et al. Digital mammography screening: weighing reduced mortality against increased overdiagnosis. Prev Med. 2011;53(3):134-40.
2. Lynge E, Ponti A, James T, Majek O, von Euler-Chelpin M, Anttila A, et al. Variation in detection of ductal carcinoma in situ during screening mammography: a survey within the International Cancer Screening Network. Eur J Cancer. 2014;50(1):185-92.
3. NHS Breast Screening Programme Annual Review 2012: NHS; 2012 [Available from: http://webarchive.nationalarchives.gov.uk/20150506150512/http:/www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp-annualreview2012.pdf.
4. Narod SA, Iqbal J, Giannakeas V, Sopik V, Sun P. Breast Cancer Mortality After a Diagnosis of Ductal Carcinoma In Situ. JAMA Oncol. 2015;1(7):888-96.
5. Bulk S, Visser O, Rozendaal L, Verheijen RH, Meijer CJ. Cervical cancer in the Netherlands 1989-1998: Decrease of squamous cell carcinoma in older women, increase of adenocarcinoma in younger women. Int J Cancer. 2005;113(6):1005-9.
6. Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E. Obstetric outcomes after conservative treatment for intraepithelial or early invasive cervical lesions: systematic review and meta-analysis. Lancet. 2006;367(9509):489-98.
7. 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.
8. Duffy SW, Michalopoulos D, Sebuodegard S, Hofvind S. Trends in aggregate cancer incidence rates in relation to screening and possible overdiagnosis: A word of caution. J Med Screen. 2014;21(1):24-9.
9. Zahl PH, Maehlen J. Overdiagnosis of breast cancer after 14 years of mammography screening. Tidsskr Nor Laegeforen. 2012;132(4):414-7.
10. Warmerdam PG, de Koning HJ, Boer R, Beemsterboer PM, Dierks ML, Swart E, et al. Quantitative estimates of the impact of sensitivity and specificity in mammographic screening in Germany. J Epidemiol Community Health. 1997;51(2):180-6.
11. Beemsterboer P. Screening for breast cancer in Catalonia. Which policy is to be preferred? Eur J Public Health. 1998:241-6.
12. Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr. 2006(36):56-65.
13. 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(11):2205-40.
14. Koleva-Kolarova RG, Zhan Z, Greuter MJ, Feenstra TL, De Bock GH. Simulation models in population breast cancer screening: A systematic review. Breast. 2015;24(4):354-63.
15. 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(7):491-9.
16. Autier P, Boniol M, Gavin A, Vatten LJ. 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.

17. Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed. 1985;20(1):79-93.
18. de Gelder R, Heijnsdijk EA, Fracheboud J, Draisma G, de Koning HJ. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. Int J Cancer. 2015;137(1):165-72.
19. de Koning HJ, Alagoz O, Schechter CB, van Ravesteyn NT. Reply to Koleva-Kolarova et al. Breast. 2016.
20. Chen J. Optima breast cancer screening policies. Masters thesis. 2013.
21. Fracheboud J. Screeningsresultaten per beoordelingseenheid 2008-2011. 2015.
22. Fracheboud J. Screeningsresultaten naar BI-RADS 2010-2013. 2015.
23. Registry TNC. Cijfersoverkanker [Webpage]. 2013 [Available from: http://www.cijfersoverkanker.nl/selecties/dataset_3/img523841be3dd9b.
24. Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klint A, et al. NORDCAN--a Nordic tool for cancer information, planning, quality control and research. Acta Oncol. 2010;49(5):725-36.
25. Bakken K, Eggen AE, Lund E. Hormone replacement therapy in Norwegian women, 1996-1997. Maturitas. 2001;40(2):131-41.
26. Bakken K, Lund E, Eggen AE. The impact of hormone replacement therapy on the incidence of breast cancer in Norway. J Clin Oncol. 2005;23(15):3636-7; author reply 7-8.
27. Francis A, Thomas J, Fallowfield L, Wallis M, Bartlett JM, Brookes C, et al. Addressing overtreatment of screen detected DCIS; the LORIS trial. Eur J Cancer. 2015;51(16):2296-303.
28. NIHR 70 day performance in initiating report Q3 2015-16. 2016.
29. Elshof LE, Tryfonidis K, Slaets L, van Leeuwen-Stok AE, Skinner VP, Dif N, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. Eur J Cancer. 2015;51(12):1497-510.
30. Worni M, Akushevich I, Greenup R, Sarma D, Ryser MD, Myers ER, et al. Trends in Treatment Patterns and Outcomes for Ductal Carcinoma In Situ. J Natl Cancer Inst. 2015;107(12):djv263.
31. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology. 2013;267(1):47-56.
32. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol. 2013;14(7):583-9.
33. Emaus MJ, Bakker MF, Peeters PH, Loo CE, Mann RM, de Jong MD, et al. MR Imaging as an Additional Screening Modality for the Detection of Breast Cancer in Women Aged 50-75 Years with Extremely Dense Breasts: The DENSE Trial Study Design. Radiology. 2015;277(2):527-37.
34. Radboud UMC; 2016 [Available from: https://www.prisma-studie.nl/.
35. de Gelder R, Draisma G, Heijnsdijk EA, de Koning HJ. Population-based mammography screening below age 50: balancing radiation-induced vs prevented breast cancer deaths. Br J Cancer. 2011;104(7):1214-20.

SUMMARY

CHAPTER 1

Breast cancer has been around for a long time. Thanks to the efforts of patients and doctors, most people became aware of the risk of breast cancer, and the treatment and detection has taken flight. Breast cancer develops in the glandular tissue of the breast or in the milkducts. There are five stages of breast cancer. The earliest stage is ductal carcinoma in situ (DCIS). In DCIS malignant cells are present in the breast, but they have not yet grown out of the normal bounds. At the next stage, invasive breast cancer, there is growth beyond the normal limits and thus also risk of metastases. Metastases in breast cancer go first to the sentinel node, a lymph node in the armpit on the same side.

Screening aims to detect breast cancer as early as possible so that the cancer is still treatable and has not spread. In the Netherlands, all women aged 50 to 74 years are invited every other year for screening. Screening is done by taking two X-rays of each breast (mammogram). These mammograms are examined by two radiologists and the GP and the patient receive a letter with the results. If there is an anomaly which requires further examination, patients are referred to the hospital. Women can also be referred because it is not possible to make a judgment based on the screening mammograms.

In Norway screening started then years later. All women from 50 to 70 years are invited every other year for mammography. Women are referred to the hospital to see if there is an abnormality.

Screening prevents breast cancer mortality, this has been shown by the first trials from Sweden. Since then several studies were conducted to evaluate whether screening is functioning properly. There are also research groups that believe that mortality reduction is moderate and that screening has many disadvantages. The main disadvantage of screening is overdiagnosis. Overdiagnosis occurs when a woman has a breast cancer diagnosis, as a result of the screening, that she would have never gotten in the absence of screening. This does not mean that she does not have breast cancer, that would be a false-positive diagnosis. It means she has a diagnosis of a disease which would never have become clinically apparent.

In this thesis, I examined the impact of overdiagnosis. Overdiagnosis is most common in DCIS. I have therefore looked at whether more DCIS is found with digital screening. I also looked at the various forms of DCIS and what the impact is on overdiagnosis. I compared the breast cancer screening program with the program for cervical cancer screening, because in cervical cancer screening the discussion on overdiagnosis is very different. I also evaluated the situation in Norway, where the program is implemented more recently, and where the incidence of breast cancer has taken a very different course in recent decades than in the Netherlands.

CHAPTER 2

In Chapter 2, I looked at the transition of the program of screening with analogue X-ray photos, printed on film, to digital X-ray photos evaluated on a computer screen. Digital screening found more DCIS. Initially, more women were recalled, but after everyone got used to the new photos, recall rates are again similar to recall rates prior to the transition.

CHAPTER 3

In Chapter 3, I compared the breast cancer screening with cervical cancer screening. The cervical screening program is older than the breast cancer screening. The cervical cancer screening detects precancerous cervical lesions (CIN), which are not yet considered to be cancer. The treatment of CIN prevents the development of cervical cancer. In cervical cancer there is also overdiagnosis. Some of the women who had a diagnosis of CIN, would have never have developed cervical cancer. This is because some of the CIN can be cleared by the body itself. Overdiagnosis of cervical cancer can result in an unnecessary loop excision or cone biopsy. These are not very invasive procedures, but a subsequent pregnancy can have rare but serious complications. Overdiagnosis in breast cancer can result in an unnecessary partial or complete mastectomy. A partial mastectomy treatment includes radiotherapy. Altogether invasive treatments. The frequency of diagnosis of cervical cancer is higher in than in breast cancer, but the consequences smaller.

CHAPTER 4

In Chapter 4, I looked at the differentiation grade of DCIS. The degree of differentiation tells us something about the degree of malignancy. In this chapter I express malignancy grade in low grade, intermediate grade, or high grade. The distribution of these grades is the same for DCIS detected by screening and for DCIS that are not detected by screening. 18% is low-grade, 31% intermediate grade, and 51% high grade. With our micro-simulation model I have calculated the degree of overdiagnosis per grade. High grade DCIS has less overdiagnosis than low-grade DCIS.

CHAPTER 5

In Chapter 5, I looked at the incidence of breast cancer in Norway. The screening program in Norway was gradually introduced between 1996 and 2005. The incidence of breast cancer in women 50-54 started to rise dramatically in 1995. This cannot be explained by the screening. Many women in Norway have used hormone replacement therapy (HRT) for menopausal symptoms in the 90s. In 2000, it became clear that this greatly increases the risk of breast cancer. In the model, I examined whether I can explain the increase in incidence with the use of HRT. I found that the expected increase in breast cancer incidence as a result of HRT use is not enough to explain the increased breast cancer incidence. I have

insufficient explanation for the increase in incidence. Other research groups had the same problem. I have solved this by adding a risk factor in the model for women. With this I could model the incidence rise. This solution is very similar to that of another group, which made use of a cohort-effect. If I look at overdiagnosis, the estimates are in line with previous estimates and are not higher than in other European countries.

CHAPTER 6

In chapter 6, I used the model described in chapter 5 to calculate the cost-effectiveness of the program. The cost-effectiveness of a program indicates how much it costs to maintain an additional year of life, after correction for decrease in the quality of life.

The cost is the cost of the screening program, but also the cost of treatment. Because you have to treat more cancer when screening for breast cancer, these costs may be higher. Because the treatment can be done at an earlier stage, the costs may be lower.

The quality of life is temporarily decreased by screening, mostly due to uncertainty about the outcome. Quality of life also decreases when receiving a breast cancer diagnosis and additional treatments, this often takes longer. The extent to which the quality of life is diminished is expressed in utilities. The utilities are multiplied by the duration of this period in order to get a factor by which the years of life are less in quality.

The Norwegian breast cancer screening program prevents 16% breast cancer mortality. This will cost 189.557 NOK per year of life adjusted for quality of life gained (QALY gained). The threshold for cost effectiveness of a program is either 20,000 to 30,000 GBP, or three times the gross national product. In Norway that threshold would be NOK 1,926,366. The program is a cost-effective measure to reduce breast cancer specific mortality.

CHAPTER 7

In chapter 7, I answer the research questions described in Chapter 1. Digital screening detects more DCIS, thus there is a higher risk of overdiagnosis. The transition to digital screening did not lead to more referrals. Cervical screening also provides a substantial amount of diagnosis, but with less impact. The distribution of low grade, intermediate grade, and high grade DCIS is the same for DCIS detected at screening, as for DCIS detected outside of screening. Of high grade DCIS fewer cases represent overdiagnosis than of low grade DCIS. In Norway, the incidence of breast cancer increased dramatically in the 90s. The use of HRT and mammography cannot explain this increase sufficiently. The percentage of overdiagnosis in Norway is similar to other countries. The Norwegian breast cancer screening program is cost effective.

What are the results of digital screening after the study period described in chapter 2? In the years following the study in Chapter 1, the referral rate of women who have their first screening examination increased from 2.8% to 6.3%. For women with a follow-up screening referral rate remained about the same at 1.7% in 2004-2010 and 2.0% in 2013.

Is the increase in breast cancer incidence in Norway unlike that in any of the other Scandinavian countries? In Finland and Sweden the incidence of breast cancer increased considerably, but nowhere as strong as in Norway.

A good number of DCIS is overdiagnosis. The possible increase in overtreatment as result of the detection of more DCIS must be addressed. Two trials are currently set up to randomize between surgery and watchful waiting for low grade DCIS. The problem with these trials is that not many women want to wait. Another strategy is to further identify tumors with a high malignant potential and those with a low malignant potential by histological characteristics.

In the future tomosynthesis will play a greater role. Tomosynthesis takes a small series of photos instead of taking just one picture of the breast, always from a slightly different angle. The result is a set of photos in which you can scroll through the breast. This technique avoids false-positive results which can occur when overprojection of breast creates a composite picture suggestive of a mass or distorsion. The sensitivity for architecture disturbances and small tumors is higher. The tomosynthesis may be performed at approximately the same dose of radiation as a conventional mammography. This means that it may also be used for screening. The implications of the use of tomosynthesis for screening will have to be studied extensively, both for the results of the program, as well as for the cost of evaluating and storing the photo-sets.

Individualized screening and treatment strategies based on personal characteristics and risk factors avoids unnecessary screening and treatment. A young woman with severe comorbidity will have little benefit from the early detection of DCIS, while an elderly woman without any comorbidity may have many years of life to gain in preventing breast cancer mortality. Risk factors may help to determine screening strategy.

SAMENVATTING

HOOFDSTUK 1

Borstkanker bestaat al heel lang. Mede dankzij de inspanningen van patiënten zijn de meeste mensen zich bewust van het risico op borstkanker en heeft de behandeling en detectie een vlucht genomen. Borstkanker ontstaat in het klierweefsel van de borst of in de melkweg-gangetjes. Er zijn vijf stadia van borstkanker. Het vroegste stadium is het ductaal carcinoom in situ (DCIS). Hierbij zijn er kwaadaardige cellen aanwezig in de borst, maar die zijn nog niet doorgegroeid buiten de normale begrenzingen. Bij het volgende stadium, het invasieve mammacarcinoom, is er doorgroei door de normale begrenzingen en hiermee ook kans op uitzaaiingen. Uitzaaiingen gaan bij borstkanker het eerst naar de schildwachtklier, een lymfklier in de oksel aan dezelfde kant.

Screening is erop gericht om borstkanker in een zo vroeg mogelijk stadium te detecteren, zodat de kanker nog goed behandelbaar is en nog niet is uitgezaaid. In Nederland worden alle vrouwen van 50 tot en met 74 jaar om het jaar uitgenodigd voor de screening. Screening gebeurt door twee foto's van elke borst te maken (mammogram). Deze foto's worden door twee radiologen beoordeeld en indien er een afwijking te zien is die nader onderzoek behoeft, krijgen de huisarts en de patiënte een brief met de uitslag. Na verwijzing naar het ziekenhuis volgt dan nader onderzoek. Vrouwen kunnen ook verwezen worden omdat het niet mogelijk is om een uitspraak te doen op basis van de foto's van de screening.

In Noorwegen zijn ze tien jaar later begonnen met screening. Alle vrouwen van 50 tot en met 70 jaar oud worden om het jaar uitgenodigd om foto's te laten maken. Ook hier geldt dat vrouwen naar het ziekenhuis worden verwezen als er een afwijking te zien is.

Dat screening op borstkanker sterfte voorkomt is gebleken uit de eerste tests die gedaan zijn in Zweden. Sindsdien zijn er meerdere studies verricht om te evalueren of de screening naar behoren functioneert. Er zijn ook onderzoeksgroepen die van mening zijn dat er niet zoveel sterfte voorkomen wordt en dat de screening vooral veel nadelen heeft. Het belangrijkste nadeel van screening is overdiagnose. Overdiagnose komt voor als een vrouw een diagnose borstkanker krijgt, als gevolg van de screening, die zij in het geval dat er helemaal geen screening aangeboden zou worden, nooit zou hebben gekregen. Het is dus niet zo dat zij geen borstkanker heeft, dat zij ten onrechte die diagnose heeft gekregen. Het is een diagnose van een ziekte waar zij nooit ziek van zou zijn geworden.

In dit proefschrift heb ik onderzocht wat de impact van overdiagnose nou eigenlijk is. Overdiagnose komt vooral voor bij DCIS. Daarom heb ik gekeken of er meer DCIS wordt gevonden met digitale screening. Ik heb ook gekeken naar de verschillende vormen van DCIS en wat de impact daarvan is op overdiagnose. Ik heb het borstkankerscreeningsprogramma vergeleken met het programma voor de screening op baarmoederhalskanker, omdat daar de discussie over overdiagnose heel anders is. En ik heb gekeken naar de situatie in Noorwegen, waar het programma korter geleden geïmplementeerd is, en waar het

voorkomen van borstkanker in de afgelopen decennia heel anders is verlopen dan in Nederland.

HOOFDSTUK 2

In hoofdstuk 2 heb ik gekeken naar de overgang van het programma van screening met analoge foto's naar het screenen met digitale foto's. Bij het screenen met digitale foto's worden er meer DCIS gevonden. Aanvankelijk werden er veel meer vrouwen verwezen, maar nadat iedereen aan de nieuwe foto's gewend was, zijn die cijfers weer vergelijkbaar met de cijfers van voor de overgang.

HOOFDSTUK 3

In hoofdstuk 3 heb ik de borstkankerscreening vergeleken met de baarmoederhalskankerscreening. De screening op baarmoederhalskanker is ouder dan de borstkankerscreening. De baarmoederhalskankerscreening detecteert voorstadia van baarmoederhalskanker (CIN), die nog niet als kanker worden beschouwd. De behandeling van zo'n voorstadium voorkomt het ontstaan van baarmoederhalskanker. Ook bij deze screening komt overdiagnose voor. Een deel van de vrouwen waarbij een CIN wordt gevonden, zou nooit baarmoederhalskanker ontwikkeld hebben. Dat komt omdat een deel van de CIN door het lichaam zelf opgeruimd worden. Overdiagnose bij baarmoederhalskanker betekent een onnodige lis-excisie of conisatie. Dit zijn op zichzelf weinig invasieve ingrepen, maar bij een eventuele zwangerschap kunnen er zeldzame, maar ernstige, complicaties optreden. Bij de borstkankerscreening betekent een overdiagnose een onnodige borstoperatie waarbij de gehele, of een deel van de borst verwijderd worden. Bij een gedeeltelijke verwijdering volgt dan ook nog radiotherapie. Al met al ingrijpende behandelingen. De frequentie van overdiagnose is hoger bij baarmoederhalskanker dan bij borstkanker, maar de gevolgen geringer.

HOOFDSTUK 4

In hoofdstuk 4 heb ik gekeken naar de differentiatiegraad van DCIS. De differentiatiegraad zegt iets over de mate van kwaadaardigheid. In dit hoofdstuk druk ik de kwaadaardigheid uit in laaggradig, gemiddeld, of hooggradig. De verdeling tussen deze graden is gelijk voor DCIS die door screening zijn gedetecteerd als voor DCIS die niet door screening zijn gedetecteerd. 18% is laaggradig, 31% gemiddeld en 51% is hooggradig. Met ons microsimulatiemodel heb ik de mate van overdiagnose uitgerekend per graad. Hooggradig DCIS heeft minder overdiagnose dan laaggradig DCIS.

HOOFDSTUK 5

In hoofdstuk 5 heb ik gekeken naar de incidentie (het voorkomen) van borstkanker in Noorwegen. Het screeningsprogramma in Noorwegen is geleidelijk geïntroduceerd tussen 1996 en 2005. In 1995 begon de incidentie van borstkanker bij vrouwen van 50-54 al dramatisch te stijgen. Dit kan niet verklaard worden door de screening. Veel vrouwen in Noorwegen hebben in de jaren 90 hormoon vervangende therapie (HRT) gebruikt voor overgangsklachten. In 2000 werd duidelijk dat dit een sterk verhoogd risico op borstkanker geeft. Met het model kijk ik de of de stijging in incidentie te verklaren is door het gebruik van HRT. Dit bleek niet het geval te zijn. Het gebruik van HRT geeft onvoldoende verklaring voor de incidentiestijging. Andere onderzoeksgroepen hadden ditzelfde probleem. Dit is opgelost door in het model een risicofactor toe te kennen aan vrouwen van een bepaalde leeftijdsgroep. Hiermee is de incidentiestijging wel te modelleren. Deze oplossing lijkt erg op die van een andere groep, die gebruik maakte van een cohort-effect. Met betrekking tot overdiagnose, liggen de schattingen in lijn met eerdere schattingen en zijn niet hoger dan in andere Europese landen.

HOOFDSTUK 6

In hoofdstuk 6 gebruik ik het model uit hoofdstuk 5 om de kosten-effectiviteit van het programma te berekenen. De kosten-effectiviteit van een programma geeft aan hoeveel het kost om een extra levensjaar te behouden, na correctie voor afname van de kwaliteit van leven.

De kosten zijn de kosten van het screeningsprogramma, maar ook de kosten voor behandeling. Omdat je meer borstkanker moet behandelen als je screent op borstkanker, zijn deze kosten mogelijk hoger. Omdat de behandeling in een vroeger stadium kan plaatsvinden, zijn de kosten mogelijk lager.

De kwaliteit van leven neemt af door de screening, meestal als gevolg van onzekerheid over de uitslag, dit duurt maar kort. De kwaliteit van leven neemt ook af door de diagnose borstkanker en de bijkomende behandelingen, dit duurt vaak langer. De mate waarin de kwaliteit van leven afneemt wordt uitgedrukt in utiliteiten. De utiliteiten worden vermenigvuldigd met de duur hiervan om een factor te krijgen waarmee de levensjaren minder zijn in kwaliteit.

Het Noorse borstkanker screeningsprogramma voorkomt 16% borstkankersterfte. Dit kost 189,557 NOK per voor kwaliteit van leven gecorrigeerd levensjaar. De drempel om een programma kosteneffectief te kunnen noemen is ofwel 20,000 tot 30,000 GBP of 3 maal het bruto nationaal product, in Noorwegen zou die drempel 1,926,366 NOK zijn. Het programma is dus kosteneffectief.

HOOFDSTUK 7

In hoofdstuk 7 geef ik antwoord op de in hoofdstuk 1 beschreven onderzoeksvragen. Digitale screening detecteert meer DCIS, daarmee is er een hogere kans op overdiagnose. Digitale screening leidt niet tot meer verwijzingen. Screening op baarmoederhalskanker levert ook een aanzienlijk aandeel overdiagnose, maar met minder consequenties. De verdeling van laaggradige, gemiddelde en hooggradige DCIS is gelijk voor DCIS gedetecteerd bij de screening, als voor DCIS gedetecteerd buiten de screening. Hooggradige DCIS geeft minder overdiagnose dan laaggradige DCIS. In Noorwegen is de incidentie van borstkanker in de jaren 90 dramatisch gestegen. Het gebruik van mammografie en HRT kunnen deze stijging onvoldoende verklaren. Het percentage overdiagnose in Noorwegen is vergelijkbaar met andere landen. Het Noorse borstkanker screeningsprogramma is kosteneffectief.

Hoe zit het nou met de digitale screening van 2010 tot 2013? In de jaren na de studie uit hoofdstuk 1 is het verwijscijfer van vrouwen die voor het eerst naar de screening komen gestegen van 2.8% naar 6.3%, voor vrouwen met een vervolgscreening is het ongeveer gelijk gebleven met 1.7% in 2004-2010 en 2.0% in 2013.

Is de incidentiestijging in Noorwegen anders dan in andere Scandinavische landen? Ook in Finland en Zweden steeg de incidentie van borstkanker flink, maar nergens zo sterk als in Noorwegen.

Een flink aantal DCIS is overdiagnose. Om te voorkomen dat veel vrouwen overbehandeld blijven worden moet de behandeling van DCIS kritisch bekeken worden. Er lopen twee trials waarin gekeken wordt of het mogelijk is om bij vrouwen met een laaggradig DCIS niet direct te behandelen, maar regelmatig de borsten te controleren met mammografie. Het lastige van deze trials is dat niet veel vrouwen daar voor voelen. Een andere strategie is om nader in kaart te brengen wat nou de histologisch kenmerken zijn van tumoren die langzaam groeien en wat de kenmerken zijn van tumoren die juist heel snel groeien.

In de toekomst zal tomosynthese een grotere rol gaan spelen. Tomosynthese maakt in plaats van 1 foto van de borst een kleine serie foto's, steeds vanuit een klein beetje andere hoek. Het resultaat is een set foto's waarmee je door de borst heen kunt bladeren. Deze techniek voorkomt fout-positieve uitslagen doordat borstklierweefsel op de foto over elkaar heen projecteert. Ook is de sensitiviteit voor architectuurverstoringen en kleine tumoren hoger. De tomosynthese kan verricht worden met ongeveer dezelfde dosis straling als een gewone foto. Dat betekent dat het ook inzetbaar is voor de screening. De implicaties van het gebruik van tomosynthese voor screening zullen uitgebreid onderzocht moeten worden, zowel voor de uitkomsten van het programma, als voor de kosten van beoordelen en opslaan van de foto-sets.

Individualisering van screenings-strategieën en behandeling op basis van persoonlijke kenmerken en risicofactoren voorkomt onnodige screening en behandeling. Een jonge

vrouw met ernstige co-morbiditeit zal weinig baat hebben bij de vroege detectie van een DCIS, terwijl een oudere vrouw zonder enige co-morbiditeit juist veel levensjaren zou kunnen winnen bij het voorkomen van borstkankersterfte. Risicofactoren kunnen helpen om screeningsstrategie te bepalen.

ANNEX

DESCRIPTION OF MISCAN BREAST

Paula van Luijt

Eveline Heijnsdijk

Harry de Koning

INTRODUCTION

This is a description of the MIcrosimulation SCreening ANalysis (MISCAN) model. The MISCAN model is a microsimulation model developed in 1985 to assess the harms and benefits of mass screening programs.

This description is specifically written for the MISCAN breast model, which is used to evaluate the Dutch Breast cancer screening program. In the past the model has been adapted to other countries.

MISCAN is a microsimulation model; it models individual life histories for all women in a given population.

The Dutch model is calibrated on Dutch demographic data to simulate a female population with the same number of women of a certain age as seen in the Dutch population. To tailor the model to a different country (population) we use demographic data of the country of interest.

The natural history of breast cancer is implemented next. We model onset of breast cancer and progression of the disease, as well as clinical detection (detection because of symptoms). For each woman in the population a time of death is modelled. This death can be the result of breast cancer or of other causes. Demographics and natural history determine the population composition and breast cancer incidence in the modeled population at any given time.

Screening is superimposed on these individual life histories. The gradual implementation and current attendance rate are modelled. Test sensitivity is put in the model to match program sensitivity.

In figure 1 the life history of a woman with breast cancer is depicted for two situations: the first without screening and the second with screening.

The model predicts:

1. Birth, based on the life table;
2. Onset of breast cancer, based on incidence data;
3. Disease progression onto clinical detection, based on data on stage distribution;
4. Death as a result of breast cancer. The model had also predicted a time of death of other causes.

In this case, in a situation without screening, death as a result of breast cancer came first. Screening advanced the breast cancer detection. The disease is now detected at an earlier stage, and the patient could be treated in time. Her breast cancer death was averted and the model now predicted a death of other causes based on the life table. She has gained life-years as a result of screening.

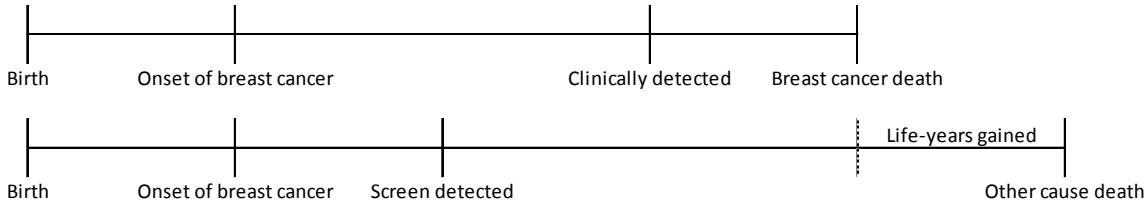


Figure 1. Life history affected by screening.

At the end the model can predict outcomes for the modelled population, these include, but are not limited to; life years; number of breast cancer diagnoses by stage, age of detection and year of detection; death by age, year and by cause of death, breast cancer or other; number of screens; number of invitations. All output can be generated for a situation with screening and for a situation without screening.

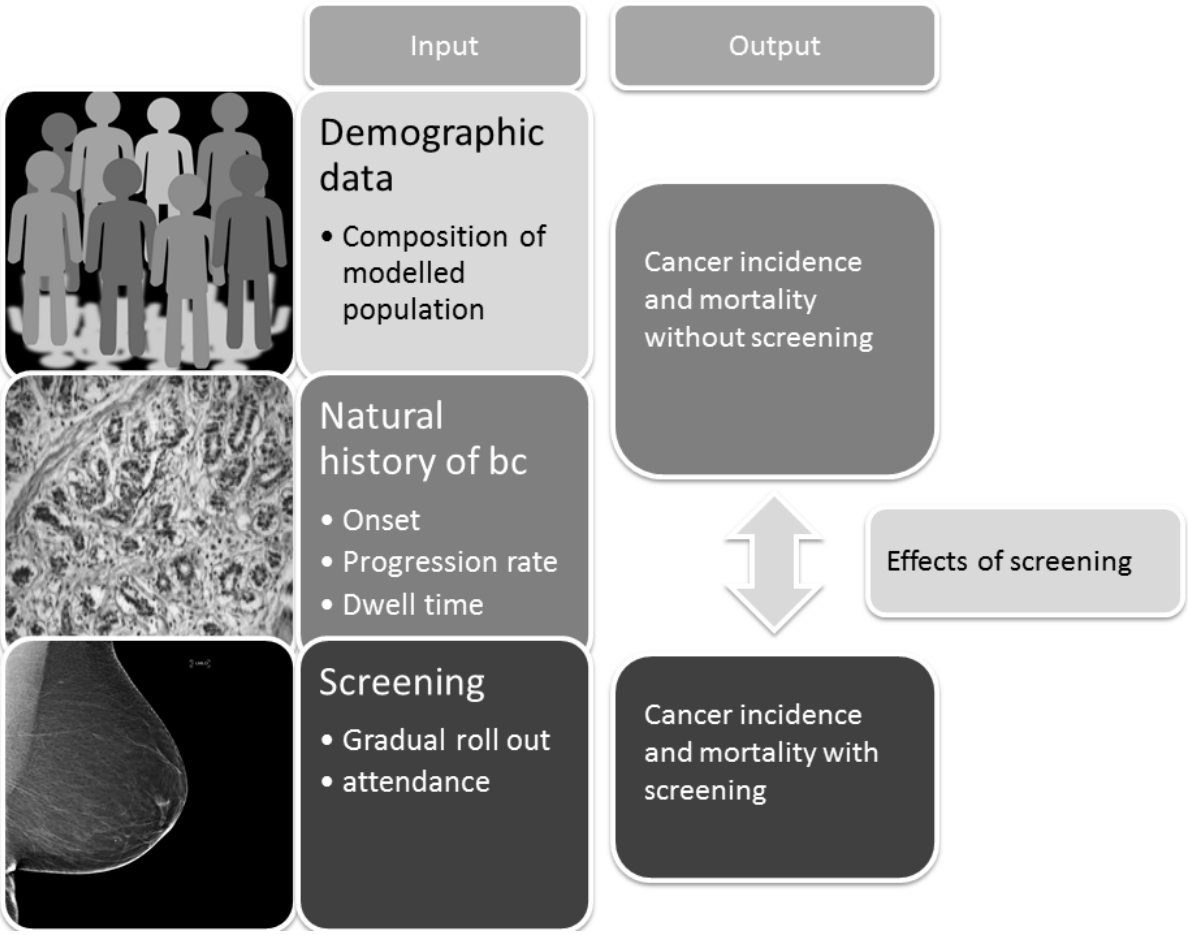


Figure 2. Graphic depiction of the input and output.

DEMOGRAPHIC DATA

Using demographic data from Statistics Netherlands we calibrate the population in the model to a population composition in a given year, by comparing the outcomes of the model with observed data. The calibration yields a birth table. Life table data from Statistics Netherlands are used to estimate all-cause mortality.(1)

The population is divided into several cohorts in order to allow for gradual implementation of screening. A given cohort has its own birth table and its own exposure to screening. Table 1 shows the distribution of births per calendar year for a cohort born between 1928 and 1989.

For each woman, a time of death from other causes (i.e. causes other than breast cancer) is generated; this time of death is independent of the breast cancer disease model. In the model, a woman's lifetime cannot exceed 100 years. The time of death from other causes is generated using a life table for women from Statistics Netherlands (Table 2).

Table 1. Birth table used in the MISCAN model for the cohort born between 1928 and 1989.

Year	Cumulative probability of birth in year
1928	0
1929	0.0063222
1934	0.069544
1939	0.13414
1944	0.20286
1949	0.29219
1954	0.38428
1959	0.48048
1964	0.58081
1969	0.68389
1974	0.77872
1979	0.85569
1984	0.92716
1989	1

Table 2. Life table used in the MISCAN model.

Age	Cumulative probability of death
0	0
5	0
10	0
15	0
20	0
25	0.0016
30	0.0034
35	0.006
40	0.0093
45	0.0141
50	0.0221
55	0.0348
60	0.0555
65	0.087
70	0.1364
75	0.2161
80	0.3424
85	0.526
90	0.7335
95	0.8962
99.9	0.9734
100	1

NATURAL HISTORY

Some of the women develop breast cancer. The probability of developing breast cancer is based on data on breast cancer incidence, growth rate, and probability of detection. To account for the increasing background incidence of breast cancer in the Netherlands, the model applies a 1.4% annual percentage change to the onset rate since 1975.(2, 3)

After having an onset women have preclinical DCIS. Four different things can happen with a preclinical DCIS: it can progress to preclinical invasive breast cancer; it can regress back to normal; it may become clinically detected; or (in the presence of screening) it may become screen-detected. All transitions through the successive disease stages are modeled using a semi-Markov process (Figure 3).

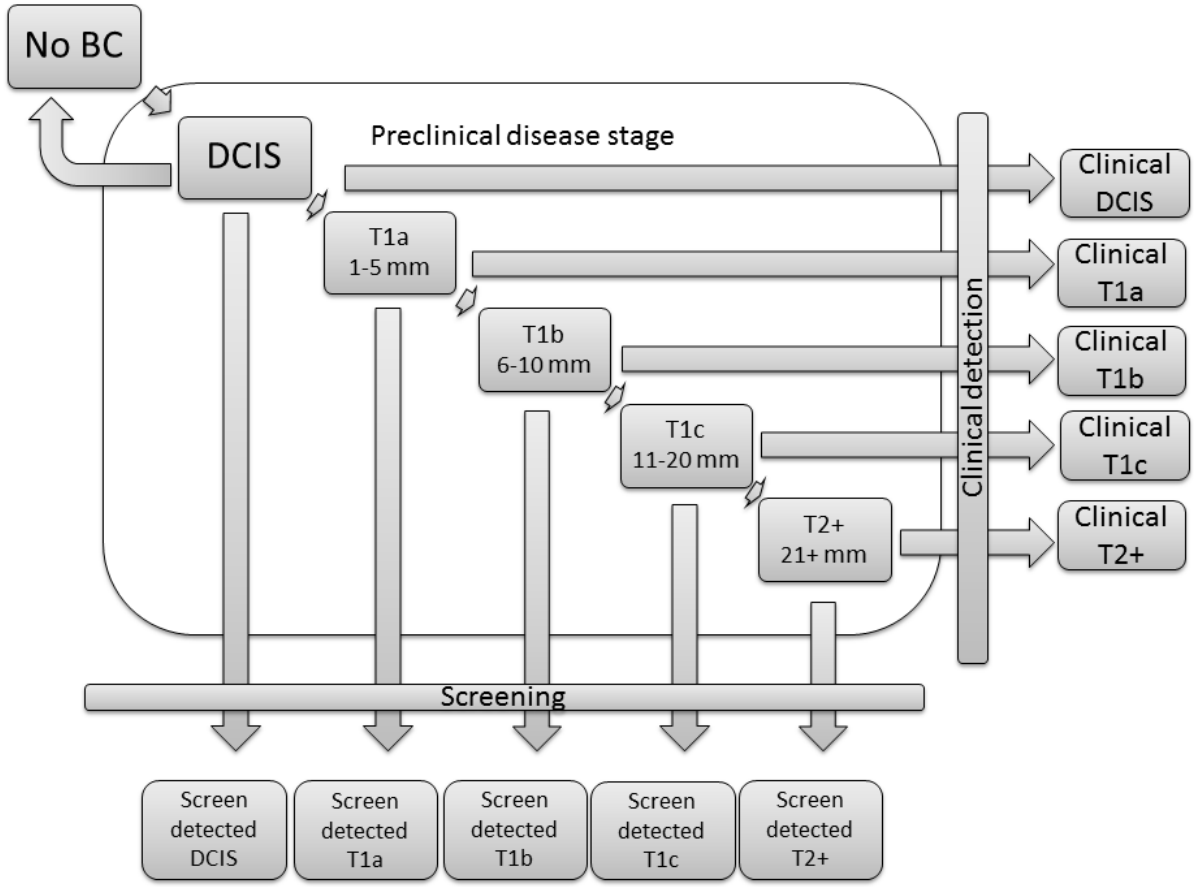


Figure 3. Schematic drawing of progression of breast cancer through the different disease stages.

The following disease stages are modelled: preclinical-, clinical- or screendetected- ; DCIS, T1a, Ta1b, T1c and T2+; each with and without lymph node involvement.

After onset, the course of the disease is modeled by :

1. The probability of progression
    a. Immediate progression
    b. Progression after dwell time
    c. Regression (only modeled for DCIS)
    d. Clinical detection
2. The time spend in a certain disease stage (dwell time)

Table 3. Probability of progression manner by disease stage.

Disease stage	Age	Clinically detected	Immediate progression	Progression after dwell time	Regression
DCIS	0	6%	76%	15%	3%
	34	6%	76%	15%	3%
	79	3%	92%	4%	1%
	100	3%	92%	4%	1%
T1a	0	2%		98%	
	35	2%		98%	
	80	1%		99%	
	100	1%		99%	
T1b	0	9%		91%	
	35	9%		91%	
	80	4%		96%	
	100	3%		97%	
T1c	0	42%		58%	
	35	42%		58%	
	80	31%		69%	
	100	26%		74%	

Progression is modeled by the probability of each disease stage to transition to the next stage and the respective dwell time. A correlation between the dwell time of the first stage of the disease and the subsequent stage of disease is modeled. Dwell times have a Weibull distribution and are age- and stage dependent.

Table 4. Mean dwell times used in MISCAN, in years by disease stage.

	Mean dwell time for a 45 year old woman (years)
DCIS	2.1
T1A	0.1
T1B	0.3
T1C	0.7
T2+	0.8

Data on incidence prior to the introduction of screening is used to calibrate onset rate, incidence, and age specific hazard ratios for the onset. These data were provided by the Eindhoven Cancer Registry from 1975-1990(4).

The incidence of breast cancer after the implementation of screening, both clinically detected and screen detected, are based on data from 1990-2008 from the Cancer Registry Netherlands(5). Incidence is specified by age, calendar year, tumor stage, and screening round (first/subsequent).

TREATMENT

Any screen detected or clinical disease stage transits directly to one of the four treated states: no adjuvant therapy (all patients receive standard care, which includes surgery and possibly radiotherapy); chemotherapy; hormonal therapy; or a combination of chemotherapy and hormone therapy. A woman in a treated state can die, either of breast cancer or from other causes.

The probability for a certain treatment is dependent on disease stage, age calendar year, and detection mode. From 1975-1990 this is based on data from the Eindhoven Cancer Registry, and no combined use of hormone and chemotherapy was assumed prior to 1990. The duration of treatment with Tamoxifen was assumed to be 5 years, and this treatment was assumed to only be administered to women with a estrogen receptor positive breast cancer. The data on treatment after 1990 was provided by the Dutch screening organization and is specified by age, tumor stage, detection mode (screendetected or clinically detected), and calendar year.

SCREENING

The gradual implementation of mass screening is mimicked using strata. The population is distributed into different strata of a certain magnitude. Each stratum is exposed to an invitation to screening at a given time, in such a way that replicates the roll out. This strategy is also used to model the extension of the program to include the women aged 70-74. At the end of roll out 100% of the women aged 50-74 are invited to screening, some since 1990 and some since later years.

The specific sensitivity for every test by age, disease stage, and calendar year, is calibrated to data on breast cancer incidence in the presence of screening (Table 5). Sensitivity in the model is for analogous mammography screening. Sensitivity is assumed to be 25% less in women under the age of 50.

Table 5. Mean test sensitivity by stage.

	Mean test sensitivity
DCIS	77%
T1A	52%
T1B	62%
T1C	90%
T2+	95%

The effects of screening on survival depend on a stage shift, which occurs due to early detection. Survival is dependent on treatment and disease stage (see also the chapter survival below).

SURVIVAL

All-cause mortality is determined by the life table. All women die at age 100 at the latest. Breast cancer survival is stage dependent. Survival rates per stage and treatment choice were modeled using international sources.(6-10)

The long-term relative survival by different treatments was previously published by de Gelder et al.(2). Relative survival is modelled by age, disease stage, and type of treatment (Tables 6 through 9). Tables are repeated here for the purpose of completeness.

Table 6. Relative survival by disease stage and age for women not receiving adjuvant treatment.

	DCIS	T1aN-	T1aN+	T1bN-	T1bN+	T1cN-	T1cN+	T2+N-	T2+N+
<30	1	0.761	0.510	0.696	0.408	0.557	0.236	0.310	0.056
40	1	0.798	0.575	0.741	0.481	0.628	0.310	0.386	0.102
50	1	0.815	0.605	0.762	0.512	0.646	0.341	0.418	0.118
60	1	0.796	0.568	0.738	0.472	0.612	0.298	0.375	0.089
70	1	0.737	0.476	0.667	0.376	0.524	0.213	0.282	0.052
≥80	1	0.678	0.383	0.597	0.279	0.435	0.128	0.189	0.016

Table 7. Relative survival by disease stage and age for women receiving hormonal adjuvant treatment.

	DCIS	T1aN-	T1aN+	T1bN-	T1bN+	T1cN-	T1cN+	T2+N-	T2+N+
<30	1	0.854	0.701	0.814	0.639	0.730	0.534	0.579	0.424
40	1	0.865	0.714	0.826	0.650	0.743	0.533	0.585	0.388
50	1	0.860	0.699	0.819	0.629	0.731	0.499	0.558	0.330
60	1	0.856	0.696	0.815	0.628	0.727	0.505	0.559	0.357
70	1	0.832	0.666	0.788	0.601	0.696	0.497	0.541	0.394
≥80	1	0.797	0.612	0.746	0.546	0.644	0.451	0.489	0.380

Table 8. Relative survival by disease stage and age for women receiving adjuvant chemotherapy treatment.

	DCIS	T1aN-	T1aN+	T1bN-	T1bN+	T1cN-	T1cN+	T2+N-	T2+N+
<30	1	0.831	0.652	0.784	0.580	0.686	0.458	0.510	0.329
40	1	0.858	0.700	0.817	0.634	0.731	0.513	0.567	0.366
50	1	0.855	0.691	0.814	0.619	0.723	0.486	0.546	0.314
60	1	0.820	0.620	0.769	0.535	0.659	0.382	0.450	0.198
70	1	0.767	0.535	0.705	0.446	0.578	0.301	0.363	0.157
≥80	1	0.720	0.464	0.649	0.373	0.509	0.241	0.294	0.144

Table 9. Relative survival by disease stage and age for women receiving both adjuvant chemotherapy and adjuvant hormonal treatment.

	DCIS	T1aN-	T1aN+	T1bN-	T1bN+	T1cN-	T1cN+	T2+N-	T2+N+
<30	1	0.905	0.806	0.879	0.765	0.824	0.697	0.726	0.626
40	1	0.912	0.814	0.887	0.772	0.833	0.697	0.730	0.602
50	1	0.890	0.765	0.859	0.711	0.790	0.611	0.655	0.481
60	1	0.872	0.729	0.835	0.669	0.757	0.559	0.608	0.428
70	1	0.851	0.702	0.811	0.645	0.730	0.552	0.592	0.461
≥80	1	0.820	0.654	0.774	0.596	0.683	0.511	0.545	0.448

CALIBRATION OF THE MODEL

Based on the observed data the model optimizes set parameters to meet the observed data in the model output. MISCAN uses the Nelder and Mead simplex ("Amoeba") multivariate minimization routine, which has been adapted for optimizing random functions. The model will be run repeatedly, and the simulation run is compared to the counts in the observations. The total deviance is the sum of the individual deviances. A convergence criterion is set by a Kendall Tau test. The required significance level of the test can be specified. An example of the results of calibration is shown in Figure 4.

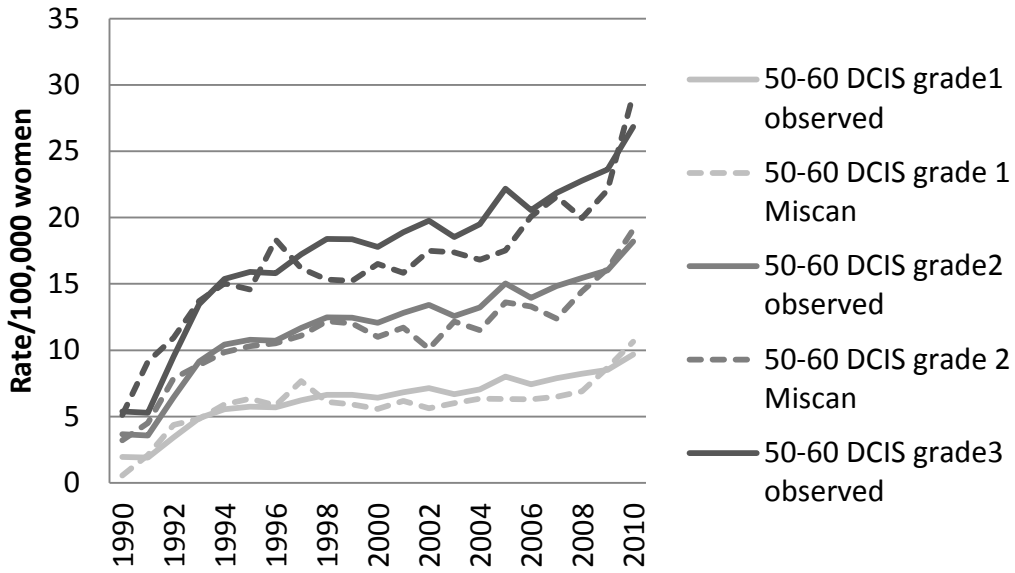


Figure 4. Example of the result of calibration. Model fit for DCIS in women aged 50-60 years. The model output is compared to the observed data.

REFERENCES

1. Statistics Norway. 2015. Available from: http://www.ssb.no/en/dode.
2. de Gelder R, Heijnsdijk EA, Fracheboud J, Draisma G, de Koning HJ. The effects of population-based mammography screening starting between age 40 and 50 in the presence of adjuvant systemic therapy. Int J Cancer. 2015;137(1):165-72.
3. Sankatsing VD, Heijnsdijk EA, van Luijt PA, van Ravesteyn NT, Fracheboud J, de Koning HJ. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int J Cancer. 2015.
4. Breast Cancer Screening IUKP. The benefits and harms of breast cancer screening: an independent review. Lancet. 2012.
5. Nederlandse kanker registratie. Available from: http://cijfersoverkanker.nl/p=4f4f666ec9d0d.
6. Schouten van der Velden AP, Van Dijck JA, Wobbes T. Variations in treatment of ductal carcinoma in situ of the breast: a population-based study in the East Netherlands. Eur J Surg Oncol. 2007;33(4):424-9.
7. Michaelson JS, Silverstein M, Sgroi D, Cheongsiatmoy JA, Taghian A, Powell S, et al. The effect of tumor size and lymph node status on breast carcinoma lethality. Cancer. 2003;98(10):2133-43.
8. NICE. NICE guide to the methods of health technology appraisal. London: 2008.
9. Tabar L, Yen MF, Vitak B, Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet. 2003;361(9367):1405-10.
10. Adami HO, Bergstrom R, Lund E, Meirik O. Absence of association between reproductive variables and the risk of breast cancer in young women in Sweden and Norway. Br J Cancer. 1990;62(1):122-6.

DANKWOORD

Ik ben dank verschuldigd aan de mensen die mij hebben begeleid bij het schrijven van dit proefschrift. In eerste plaats Eveline, die altijd beschikbaar was en klaar stond om weer de volgende stomme fout uit het model te halen of mee te denken over waarom iets niet lukte. Alles wat ik ter beoordeling opstuurde was altijd binnen de kortste keren weer terug en voorzien van duidelijk en opbouwend commentaar.

In de tweede plaats natuurlijk Harry. Ondanks zijn zeer drukke agenda was er altijd op korte termijn een afspraak te regelen en ik waardeer je persoonlijke betrokkenheid bij mijn onderzoek. Ik heb het altijd bewonderenswaardig gevonden hoe snel je uit een tabel een opmerkelijk getal kan vinden en daarmee het probleem van het hele onderzoek kan duiden.

Ik wil ook graag Nicolien bedanken omdat ze altijd te vinden was voor een praatje en altijd meedacht over model problematiek. Ik bewonder je zeer pragmatische en logische kijk op de dingen.

Arry, je bent onmisbaar voor het LETB en voor de screensectie. Je hebt zo veel ervaring in de LETB. Je zat ons achter de broek voor onze publicaties en voor input in de LETB offerte. Altijd was je bereid om me te helpen en bovendien was je altijd te vinden was voor een gezellig praatje. Bedankt voor alle moeite die je hebt gedaan om alles op tijd bij de commissie te krijgen.

Jacques, ik wil je bedanken voor het geduld waarmee je mee hebt geholpen de LETB data te leren begrijpen. Ik waardeerde ook de discussies die zich uitstrekten tot meer dan alleen het LETB.

Mijn carrière heeft niet een heel logisch pad gevolgd en ik wil de mensen met nadruk bedanken die mij toch de kans hebben gegeven om een en ander voor elkaar te krijgen. Harry omdat ik de promotieplek kreeg op een moment dat ik uit de kliniek wegging en mijn carrière op een andere manier vorm moest gaan geven. Maar ook Tadek, omdat die mij een plek gaf op het moment dat ik de kliniek weer terug in wilde en daarnaast ook nog de mogelijkheid bood om deze promotie af te ronden naast mijn klinische werkzaamheden.

Fleur bedankt omdat je, naast al het andere, ook mijn paranimf wil zijn.

CURRICULUM VITAE

Paula van Luijt was born on May 23rd 1980 in The Hague, the Netherlands. She graduated at the Maerlant lyceum 1998. After this she studied medicine at the Erasmus MC in Rotterdam. After obtaining her masters in 2003, she took part in the board of her students association. In 2006 she became a certified MD. From September 2006 to September 2008 she worked in the Ikazia ziekenhuis in Rotterdam as a resident in obstetrics and gynaecology. From October 2008 to February 2009 she worked at the Maasstad ziekenhuis in Rotterdam, also as a resident in obstetrics and gynaecology. In February 2009 she began her training as a radiology resident in the Leids Universitair medisch centrum in Leiden. She interrupted her training in 2011 to start on her PhD at the department of Public Health at the Erasmus MC in Rotterdam. After two years she resumed her radiology training at the Albert Schweitzer ziekenhuis in Dordrecht, meanwhile completing the work on the PhD. She hopes to finish her radiology training in April 2018.

Paula van Luijt werd geboren op 23 mei 1980 in Den Haag, Nederland. Ze haalde haar eindexamen op het Maerlant Lyceum in Den Haag in 1998. Hierna ging ze geneeskunde studeren op de Erasmus universiteit Rotterdam. In 2003 haalde ze haar doctoraal. Daarna heeft ze een jaar zitting genomen in het bestuur van haar studenten vereniging. In 2006 behaalde ze haar arts-diploma. Van september 2006 tot september 2008 werkte ze als arts-assistent niet in opleiding bij de afdeling gynaecologie en verloskunde van het Ikazia ziekenhuis. Aansluitend heeft ze tot februari 2009 in het Maasstad ziekenhuis gewerkt als arts-assistent niet in opleiding bij de afdeling gynaecologie en verloskunde. In februari 2009 begon ze met haar opleiding radiologie in het Leids Universitair Medisch Centrum in Leiden. Ze heeft de opleiding in februari 2011 onderbroken om te gaan werken aan haar promotie bij de afdeling Maatschappelijke gezondheidszorg in het Erasmus MC in Rotterdam. Na twee jaar heeft ze haar opleiding weer opgepakt in het Albert Schweitzer Ziekenhuis in Dordrecht, onderwijl het werk aan de promotie afmakend. In april 2018 hoopt ze klaar te zijn met de opleiding.

LIST OF PUBLICATIONS

1. van Luijt PA, Dijksterhuis MG. [Diagnostic image. A woman in childbed with itch and vesicles] Een kraamvrouw met jeuk en blaasjes. Ned Tijdschr Geneeskd. 2009;153:B356.

2. van Luijt PA, Fracheboud J, Heijnsdijk EA, den Heeten GJ, de Koning HJ, National Evaluation Team for Breast Cancer Screening in Netherlands Study G. Nation-wide data on screening performance during the transition to digital mammography: observations in 6 million screens. Eur J Cancer. 2013;49(16):3517-25.

3. Sankatsing VD, Heijnsdijk EA, van Luijt PA, van Ravesteyn NT, Fracheboud J, de Koning HJ. Cost-effectiveness of digital mammography screening before the age of 50 in The Netherlands. Int J Cancer. 2015.

4. van Luijt PA, Heijnsdijk EA, Fracheboud J, Overbeek LI, Broeders MJ, Wesseling J, et al. The distribution of ductal carcinoma in situ (DCIS) grade in 4232 women and its impact on overdiagnosis in breast cancer screening. Breast Cancer Res. 2016;18(1):47.

5. van Luijt PA, Rozemeijer K, Naber SK, Heijnsdijk EA, van Rosmalen J, van Ballegooijen M, et al. The role of pre-invasive disease in overdiagnosis: A microsimulation study comparing mass screening for breast cancer and cervical cancer. J Med Screen. 2016.

6. van Luijt PA, Heijnsdijk, E.A.M., van Ravesteyn, N.T., Hofvind, S., de Koning, H.J. Breast cancer incidence trends in Norway and estimates of overdiagnosis. J Med Screen. 2016; in press.

7. van Luijt PA, Heijnsdijk EA, de Koning HJ. Cost-effectiveness of the Norwegian breast cancer screening program. 2016. Submitted.

PHD PORTFOLIO

	Year	ECTS	Hours
Courses			
Planning and evaluation of screening	2011	1.4	
Biostatistical Methods I: Basic Principles	2011	5.7	
Biostatistical Methods II: Popular Regression Models	2011	4.3	
Courses for the quantitative researcher	2012	1.4	
Repeated measurements in Clinical Studies	2012	1.4	
Literature search	2012		2.5
Advanced literature search	2012		2.5
Endnote	2012		2.5
Masterclasses summer course 2011	2011		4
Writing course	2012	4	
Total courses		18.2	11.5
Conferences and Seminars			
Lunch seminars (60 hours)	2011/2012	2	
Oral presentation seminar	2011	2	
Club Meth meetings	2011/2012		44
Attendance EBCC08	2012	1	
Oral presentation EBCC08	2012	1	
Attendance "Preventing overdiagnosis, winding back the harm of too much medicine"	2013	1	
Oral presentation "Preventing overdiagnosis, winding back the harm of too much medicine"	2013	1	
Attendance "Preventing overdiagnosis, winding back the harm of too much medicine"	2014	1	
Oral presentation "Preventing overdiagnosis, winding back the harm of too much medicine"	2014	1	
Total conferences and seminars		10	44
Teaching			
Bachelor thesis 2011	2011		60
Supervise intern	2011		50
Community Project 2012	2012		12
Skills education, medication safety	2012		14
Total teaching		0	136

Total ECTS	**28.2**	**191.5**
Hours to ECTS (=hours/28)	**6.8**	
Total PhD period	**34**	