The Research Council of Norway

# Research-based evaluation of the Norwegian Breast Cancer Screening Program

Final report

Evaluation
Division for Society and Health

The Research Council of Norway

# Research-based evaluation of the Norwegian Breast Cancer Screening Program

**Final report**

Evaluation
Division for Society and Health

The Research Council of Norway
P.O.Box 564
NO-1327 Lysaker

Telephone: +47 22 03 70 00
Telefax: +47 22 03 70 01
post@rcn.no
www.rcn.no/english

Oslo, May 2015
ISBN 978-82-12-03424-2 (print)
ISBN 978-82-12-03425-9 (pdf)

The report can be ordered and downloaded at
www.forskningsradet.no/publikasjoner

Number of copies: 200
Printing: 07 Gruppen/The Research Council of Norway
Design cover: Design et cetera

# Preface from the Research Council of Norway

The Research Council of Norway has conducted a research-based evaluation of the Norwegian Breast Cancer Screening Program on behalf of the Ministry of Health and Care Services. The total budget for the evaluation was NOK 18.2 million. The evaluation period lasted from 2007 to May 2015.

The objective of the research-based evaluation was to investigate whether the Norwegian Breast Cancer Screening Program fulfils its intentions and purpose. In this program, all women aged 50-69 years are invited to undertake a mammography screening every other year. In the evaluation, special focus was placed on whether the Screening Program has achieved its primary target of a 30 per cent reduction in breast cancer-related mortality among women invited to take part in the screening. The evaluation also gives information about the extent of overdiagnosis, the women's experience with mammography screening and of the cost-effectiveness.

Research-based evaluation projects have to fulfil the same requirement of high scientific quality as ordinary research projects. However, in research-based evaluations, there is a need for the commissioning authority to influence the scope of the projects, and the Research Council of Norway has therefore also had an important role to secure the necessary distance between the Ministry of Health and Care Services and the researchers.

The Research Council appointed a steering committee to ensure the scientific follow-up of the research-based evaluation of the Norwegian Breast Cancer Screening Program. The steering committee has consisted of the following:

- Professor Roar Johnsen (chair), Norwegian University of Science and Technology (15.5.07 – 31.5.15)
- Professor Jan P.A. Baak, Stavanger University Hospital (15.5.07 – 30.1.14)
- Associate Professor Anne Wenche Emblem, University of Agder (15.5.07 – 31.5.15)
- Professor Lars Holmberg, King's Collage London, Regional Oncologic Centre, Sweden (15.5.07 – 31.5.15)
- Deputy CEO Tone Ikdahl, Akershus University Hospital (15.5.07 – 31.5.15)
- Professor Niels Keiding, Copenhagen University, Denmark (15.5.07 – 31.5.15)
- Director (in pension) Berit Mørland, Norwegian Knowledge Centre for Health Services (15.5.07 – 31.5.15)
- General Practitioner Marte Walstad, Ranheim Medical Centre (10.9.07 – 31.5.15)
- Anne-Grethe Rørstad (observer), Breast Cancer Association (27.2.15 – 31.5.15)

Researcher Signe Opdahl has been engaged to draft the evaluation report on behalf of the steering committee and has worked in close cooperation with the committee on the analysis and compilation of the findings in the evaluation.

The Research Council of Norway wants to sincerely thank the steering committee for the immense work they have put into the evaluation and the report, and for their endurance and never failing engagement during the time the evaluation was on halt due to

different difficulties related to access to data. The work of Signe Opdahl is greatly acknowledged.

The Research Council of Norway also want to thank the Cancer Registry of Norway for the good collaboration and all the researchers involved in the evaluation projects for their work.

This report finalizes the evaluation assignment as given by the Ministry of Health and Care Services to the Research Council of Norway.

Oslo, May 2015

Mari K. Nes
Director
Society and Health

# Preface from the Steering Committee

The Research Council of Norway appointed in May 2007, a steering committee to conduct a research-based evaluation of the Norwegian Breast Cancer Screening Program (NBCSP). The competence of the members includes epidemiology, medical statistics, public health, oncology, health economics, pathology, and family medicine.
Detailed information on competence and conflict of interests may be found on http://www.forskningsradet.no/prognett-mammografi/Styringsgruppen/1226994052806.
In addition, an observer from the Breast Cancer Association has joined the committee through all meetings and conferences. In 2013, MD and researcher Signe Opdahl was engaged to draft the evaluation report. She has worked in close cooperation with the committee on reviewing and analyzing the current literature, and in analyzing and reviewing the publications and reports provided in the evaluation.

During the first meeting, the committee discussed thoroughly whether a new evaluation based on observational studies could add new knowledge on benefits and harms on breast cancer screening. The results from 11 conducted comprehensive randomized controlled trials had not generated mutual conclusions on the efficiency of breast cancer mass screening. The two main topics for the discrepancies are the efficiency on breast cancer mortality and on the potential major harm of screening: overdiagnosis. Implementation of new technology has changed the diagnostic accuracy from the late sixties through the nineties, which was also an argument to perform the intended evaluation. The committee recognized that an evaluation of the NBCSP based on observational studies, could hardly add anything to the assessment of the efficiency of breast cancer screening. The NBCSP has, however, some characteristics that could be important to evaluate. First, the screening program was implemented stepwise, although not randomly, with four pilot counties in 1995-97, and then the other 15 counties through 2005, where all women aged 50-69 were invited. The stepwise inclusion make it possible to compare exposed (invited) to non-exposed (not invited). Second, a special registry for NBCSP could provide all information on the date of exposure, attendance, and the results of all mammograms taken. Besides, since all inhabitants of Norway have a unique personal identification number, it was possible to use, for all purposes, individual data. Third, national registries such as the Norwegian Causes of Deaths Registry, the Norwegian Cancer Registry, the Medical Birth Registry of Norway, Statistics Norway (changes of address, civil status, education etc.) are recognized as high quality registries. Based on this knowledge, the committee accepted the appointment, still well aware of the challenges of the non-randomized implementation (county differences in breast cancer incidence and mortality), the short follow-up, potential effects of non-program screening, the period of rather extensive use of hormone replacement therapy, and the challenge of not unravelling effects of screening and improved therapy.

The Ministry of Health and Care Services expected the evaluation to address changes in breast cancer mortality and stages of breast cancer, interval cancer and overdiagnoses, and the women's perspective on the program, and to include a health economic evaluation.

The invitation to apply for project founding was announced both internationally and nationally, however surprising few applications were received. The committee decided to have at least two different project groups to assess breast cancer

mortality, overdiagnosis and health economic analyses, and that these groups should have the same sets of data from a common project database.

Early in the process, but after the Research council had contracted the research groups, it became evident that the screening program database was mainly constructed to serve as a patient administrative tool. To fulfil the requirements for a quality assured project database a considerable reorganization had to be implemented. This postponed the start of the evaluation with at least one year. The Norwegian Data Inspectorate had previously imposed the Cancer Registry to obtain informed consent from participating women to store information on negative mammograms. For several screening rounds consents were missing and the Inspectorate could not approve use of the data and required stored data deleted. Such an action would have distorted the evaluation, eventually the problem was solved, but with additional loss of time. Meanwhile, the project groups had appointed researchers to carry out the projects with no access to individual data linked to other registries. Consequently, when the project database was readily available, some of the research groups' capacity was already used, and there was no supplementary funding available. These obstacles have hampered the evaluation process, and to some extent influenced the source material and the evaluation results.

There have been ups and downs in the work during the eight years. The downs are accounted for above; the ups have been the collaboration with the project groups, the communication with professionals at the Cancer Registry, and last but not least, the outstanding administration from the Research Council, represented by Henrietta Blankson. None of project groups resigned during the eight years. The Committee is impressed by and grateful for their never-ending enthusiasm in creating and communicating good research, repeatedly answering questions and their engaged participation on seminars. The evaluation results rest heavily on their contributions.

Signe Opdahls' first task was to sum up the potential methodological challenges in assessing the central aims in the evaluation, especially breast cancer mortality reduction and overdiagnosis. Her methodological skills and immense precision in her presentations engaged the committee in the final rounds of discussions of the evaluation results. She has challenged the members of the committee to provide their best, and the committee is highly grateful for her contributions and hard work.

Have the eight years of evaluation been worthwhile? Despite several obstacles, some difficult to understand, the evaluation process has brought about a project database of high quality for new studies and further evaluation of the program, and a meticulous review of the literature on methodological challenges in evaluation based on observation studies. The report could have been more precise in some of the estimates, but considering the many challenges in evaluating a screening program by observational studies, the committee hopes that the results will contribute to knowledge based decisions on breast cancer screening.

The members of the steering committee express their sincere thanks to the Norwegian Research Council for the assignment and for excellent support during the evaluation period.

# Contents

# Summary


## *Breast cancer in Norway*

Breast cancer is the most common form of cancer among women in Norway and constitutes 20% of all cancers in Norwegian women. In 2012, there were 2956 new cases of invasive breast cancers for women of all ages combined. Based on the incidence in 2012, it is estimated that the risk of getting a breast cancer diagnosis before age 85 years is 10.5% for Norwegian women. Ductal carcinoma in situ (DCIS) is a pre-invasive neoplastic lesion in the breast and is considered a non-obligate precursor of invasive breast cancer. In the prescreening period (before 1995), DCIS represented about 5% of all malignant breast tumors, compared to 10-20% in populations offered mammography screening. In the age group 50-69 years, approximately 67% of breast cancers (invasive breast cancer and DCIS combined) are now diagnosed through the Norwegian Breast Cancer Screening Program.

In the early 1990s, there were approximately 660 deaths from breast cancer per year among women 50 years and older in Norway, compared to 588 deaths from breast cancer in 2012. The total number of deaths among women in this age group was 21 178, and deaths from breast cancer thus constituted 2.6% of all deaths among women 50 years or older in 2012.

## *The Norwegian Breast Cancer Screening Program (NBCSP)*

The NBCSP invites all women aged 50-69 years for biennial mammography screening. Attendance is 76% in each screening round, whereas 83% of the invited women attend at least one examination. The screening program was implemented stepwise, although not randomly, with four pilot counties starting in 1995-97, and the remaining 15 counties gradually entering until 2005. The stepwise inclusion enables comparison between exposed (invited) and unexposed (not invited). The NBCSP database could provide all information to researchers on the date of invitation, attendance, and the results of all mammograms taken. Since all inhabitants of Norway have a unique personal identification number, it was possible to use individual data and to link information from national registries such as the Norwegian Causes of Deaths Registry, the Cancer Registry of Norway, the Medical Birth Registry of Norway and Statistics Norway, all of which are recognized as high quality registries.

## *The evaluation assignment*

In 2006, the Ministry of Health and Care Services charged the Research Council of Norway with responsibility for conducting a research-based evaluation of the Norwegian Breast Cancer Screening Program. The Research Council invited researchers to forward project applications and Norwegian as well as foreign institutions could be project owners. An external international expert committee evaluated the scientific quality of all eleven applications. The steering committee accepted seven applications and found it important that the prioritized research questions were investigated from different perspectives and with different methods to provide a broad evidence basis. The steering group thus decided to invite at least two different research groups to evaluate mortality reduction, overdiagnosis and health economy. The funded sub-projects covered the

following themes: Changes in breast cancer mortality, interval cancer occurrence, extent of overdiagnosis, women's experiences, economic evaluation and cost-effectiveness analysis.

## *Results*

The basis for this evaluation has been observational studies of the Norwegian Breast Cancer Screening Program published from 2008 to 2014, most of which used data from the same quality controlled project database. All studies providing an estimate of program effectiveness indicated a reduction in breast cancer mortality for women invited to mammography screening, but with variation in the reduction and in statistical precision. Use of individual data on invitation date, diagnosis and death, long follow-up and detailed adjustment for time trends and regional differences were considered as important factors in validity assessment. The potential impact of concurrent widespread use of opportunistic screening, centralization of breast cancer management and improved treatment regimens are remaining limitations in most studies. A summary measure across the studies of reduction in breast cancer mortality attributable to the implementation of the NBCSP, compared to a situation with no screening program, is considered to be in the range 20-30% for women aged 50-79 years.

In studies of overdiagnosis following NBCSP implementation, the variation in study design and analytical approach, and also in the choice of denominator for overdiagnosis, resulted in a large range of estimates. Again, individual data, long follow-up and assumptions on breast cancer incidence in the absence of screening were key factors in estimation of the excess incidence in screening and the post-screening drop in incidence. We consider the most reliable estimates of overdiagnosis of invasive breast cancer and DCIS combined, for women aged 50-79 years compared to a situation without screening, to be within the range 15-25%. For women aged 50-79 years in a situation with screening, we consider the corresponding estimates to be within the range 15-20%.

The total costs of one screening round was estimated to NOK 574 million, or NOK 1389 per woman attending screening in 2012, including costs of screening examination, recall examinations and indirect costs, but not costs associated with running the mammography program as such. The estimated 10-year treatment costs for breast cancer in 2008 were NOK 356 000 for treatment of one patient. These costs were used to estimate cost-effectiveness of NOK 190 000 to 479 000 per quality-adjusted life-year gained by screening, and depended mainly on the number of breast cancer deaths prevented.

Interval cancers were estimated to comprise approximately 25% of cancers among women attending screening, including both invasive cancer and DCIS. One third of interval cancers may be tumors missed at previous screening (false negative), whereas two thirds may be true interval cancers. Diagnostic delay due to negative screening mammography exists, but the extent is not known. Women with interval cancers generally remain confident in the screening program.

Among women attending all 10 screening invitations, approximately 20% were estimated to experience at least one recall for further examinations due to false positive results. The increase in mental distress following a recall examination declines over time, but may recur at subsequent screening invitations or examinations. Women attending screening express greater concern for interval cancers than for false positive results. The primary motivation for attendance is to get a confirmation that they do not have breast cancer. Continued participation increases the feeling of routine and being part of a

production line. Pain and emotional distress is perceived as less dominating in subsequent screening rounds compared to the first participation

## *Balancing benefits and harms*

To enable a more direct comparison of the results summarized above, we have applied the results in absolute numbers to an expected cohort of 10 000 women aged 50 years who are invited for 10 screening rounds and followed for their remaining lifetime. We assume that 76% of the invited women attend all 10 screening examinations, whereas the remaining 24% never attend. All calculations were made for invasive cancer and DCIS combined. There is considerable uncertainty in these numbers, reflecting both the uncertainty in the included studies and in the assumptions made in the calculations.

*Breast cancer deaths prevented*

The expected number of breast cancer deaths prevented per 10 000 invited women was approximately 27.

*Overdiagnosis*

The expected number of screening-detected breast cancers for this cohort was approximately 377. Approximately 142 of these women would be expected to get a diagnosis of invasive breast cancer or DCIS that would not be detected in the woman's lifetime in a situation without screening.

*Balancing breast cancer deaths prevented and breast cancer overdiagnoses*

Combined, we consider the numbers presented above to be compatible with approximately five overdiagnosed breast cancers per breast cancer death prevented.

*False positive and false negative mammograms*

Of the 10 000 invited, approximately 1520 would experience at least one recall examination with the conclusion of a false positive mammogram. The majority of these women will be cleared from cancer suspicion after a second mammography or an ultrasound examination, whereas 310 attending women would be expected to be cleared for cancer suspicion only after invasive tests such as cytology or biopsy.

The expected number of interval cancers was approximately 127, and among these, approximately 42 would have a false negative mammogram at the preceding screening.

*True negative mammograms*

The remaining 5576 of the screening attendants would be expected to have only true negative mammograms at all 10 screening rounds.

## *Conclusion*

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.

From a societal perspective, recognizing the uncertainty of the estimates, the cost-effectiveness of the program seems to be within the range of what Norwegian Health Authorities define as acceptable for health services. On the individual level, however,

each invited woman has to weigh the information on potential benefits and harms based on her own values, health and life situation when deciding on whether or not to attend the program.

# Norsk oppsummering

Dette er en oppsummering av de viktigste punktene i evalueringen av Mammografi-programmet med hovedvekt på utfordringer, resultater og diskusjon av disse. Vi henviser til hovedrapporten på engelsk for mer detaljerte beskrivelser og referanser.

**Brystkreft i Norge**

*Infiltrerende brystkreft*
Brystkreft er den vanligste kreftformen blant kvinner i Norge. I 2012 ble det diagnostisert 2 956 nye tilfeller av infiltrerende brystkreft. For norske kvinner er det forventet at 10,5% vil få påvist brystkreft innen 85 års alder. Insidensratene av brystkreft (forekomst per 100 000 person-år) har forandret seg de siste tiårene. Frem til 1980-årene økte brystkreftratene jevnt blant kvinner i alder 40 år og eldre. Fra midten av 1990-årene har det vært en betydelig økning blant kvinner 50-69 år, etterfulgt av en liten nedgang utover 2000-tallet. I samme periode har det vært en stabilisering etterfulgt av en svak nedgang blant kvinner over 70 år.

Brystkreft er en heterogen sykdom med svært varierende klinisk sykdomsbilde. Behandling av brystkreft har utviklet seg betydelig siden 1980-årene, og består i dag av kirurgi, strålebehandling, cellegift og antihormonell behandling. Behandling av den enkelte pasient bestemmes blant annet av sykdommens utbredelse ved diagnosetidspunktet (stadium), histologisk vurdering av svulsten (histologisk type og grad) og molekylære markører som hormonreseptorer.

*Ikke-infiltrerende brystkreft*
Før Mammografiprogrammet utgjorde ikke-infiltrerende brystkreft (DCIS – duktalt carsinoma in situ) rundt 5% av all brystkreft, sammenliknet med rundt 13% blant kvinner som inviteres til Mammografiprogrammet. DCIS er også en heterogen tilstand, og det er usikkert hvor stor andel som utvikler seg til infiltrerende brystkreft. I små studier av kvinner med DCIS oppdaget utenfor screening og som har blitt fulgt uten behandling, er det antydet at rundt 40% kan utvikle seg til infiltrerende brystkreft i løpet av 30 års oppfølging. DCIS behandles i dag rutinemessig med kirurgi og strålebehandling.

*Dødelighet av brystkreft*
Dødeligheten av brystkreft øker med stigende alder. I aldergruppene opp til 70 år var dødeligheten stabil fram til midten av 1990-årene hvoretter den har falt. For kvinner 70 år og eldre steg dødeligheten frem til midten av 1990-årene hvoretter den også har falt. Antall dødsfall av brystkreft blant kvinner 50 år og eldre falt i samme periode fra rundt 660 per år til 588 i 2012. Rundt 2,6% av alle dødsfall blant kvinner 50 år og eldre skyldes brystkreft.

*Risikofaktorer for brystkreft*
Det er mange kjente risikofaktorer for brystkreft knyttet til reproduksjon og livsstil. Spesielt viktig i denne sammenhengen er bruk av hormoner i overgangsalderen som har vist seg å øke risikoen for brystkreft. Bruk av slike hormoner økte vesentlig samtidig med innføringen av Mammografiprogrammet. Sammenliknet med slutten av 1980-årene var

bruken mer enn firedoblet 10 år senere. Etter at det i 2002 kom resultater fra kontrollerte forsøk som viste økt risiko for både brystkreft og hjerte-kar-sykdom blant kvinner som brukte kombinasjonsbehandling (østrogen og progesteron), falt bruken igjen mot slutten av 2009. Studier fra den norske kohorten «Kvinner og kreft» har vist at i aldersgruppen 48-62 år brukte 38% i 2002 slike hormoner. De som brukte hormonbehandling som inneholdt både østrogener og progesteroner, hadde i samme studie en mer enn doblet risiko for brystkreft sammenliknet med de som aldri hadde brukt slike hormoner. Blant kvinner som deltok i Mammografiprogrammet mellom 1996 og 2004 oppga 43% at de hadde brukt hormonbehandling.

## Prinsipper ved mammografiscreening for brystkreft

Screening innebærer å undersøke symptomfrie personer for å oppdage sykdom. Verdens helseorganisasjon har beskrevet 10 prinsipper som indikerer om screening for en sykdom kan være nyttig. Hovedprinsippene er at sykdommen må ha en latent fase hvor den kan diagnostiseres før den har gitt symptomer, og at tidlig diagnose etterfulgt av behandling bedrer prognosen (forløpet) sammenliknet med tilsvarende sykdom som blir diagnostisert etter at sykdommen har gitt symptomer. Dette krever at sykdommens naturlige forløp er kjent. Siden brystkreft er en svært heterogen sykdom, er det vanskelig å kjenne forløpet for alle varianter. Varigheten av latensfasen og varigheten av svulstenes ledetid (tiden fra svulsten blir oppdaget på screening til den ellers ville gitt symptomer) er derfor i praksis ikke mulig å måle.

Den samlede nytten av et screeningprogram bestemmes av balansen mellom fordeler og ulemper. Kvinner som får oppdaget brystkreft på screening (de sanne positive) vil kunne oppleve fordelene ved screening gjennom tidlig diagnose og behandling med redusert risiko for å dø av sykdommen. Samtidig vil screening også oppdage svulster som ikke ville ha gitt symptomer i kvinnens levetid og derfor ikke ville blitt oppdaget uten screening (overdiagnostikk). Det er ikke mulig å vite ved diagnosetidspunktet om en svulst er overdiagnostisert eller ikke. Dermed vil også disse kvinnene bli behandlet og oppleve bivirkninger av behandling i tillegg til belastningen av en brystkreftdiagnose. Kvinner som ikke har brystkreft og blir vurdert som friske ved mammografi (de sanne negative) får gjennom screening forsikring om at de ikke har brystkreft. For kvinner med mammografibilder som gir mistanke om kreft, men som viser seg ikke å ha brystkreft ved etterundersøkelser (de falske positive), vil etterundersøkelser og ventetid kunne medføre psykisk og fysisk belastning. Kvinner som har brystkreft uten at dette blir fanget opp på screening (de falske negative) vil kunne oppleve falsk trygghet med risiko for forsinket diagnose og behandling.

## Det norske Mammografiprogrammet

I november 1985 oppnevnte Helsedirektøren en arbeidsgruppe som gjennomførte en bredt anlagt evaluering av mulighetene og konsekvensene av en norsk screening for bryst kreft, og som anbefalte oppstart med mammografiscreening for kvinner 50-74 år hvert annet år. I 1994-95 bevilget Stortinget midler til oppstart av et prøveprosjekt i fire fylker (Oslo, Akershus, Hordaland og Rogaland) for kvinner 50-69 år med mammografi hvert annet år. I 1997 ble det bestemt at tilbudet skulle utvides til de øvrige fylker. Ordningen ble gjennomført trinnvis og i 2005 var alle kvinner i målgruppen i alle fylker invitert. Hovedmålet for Mammografiprogrammet har vært at det skulle føre til 30% redusert dødelighet av brystkreft blant inviterte. Screeningen foregår ved 16 brystdiagnostiske

sentre som har ansvar for diagnostikk og behandling av all brystkreft. Ved hvert senter er det etablert flerfaglige team med røntgenleger, radiografer, patologer, sykepleiere, kirurger og kreftleger. Rundt 60% av de inviterte møter regelmessig, mens 83% av alle inviterte møter til minst én undersøkelse. Ved hver screeningrunde møter omtrent 76% av de inviterte. For perioden 1996-2007 utgjorde brystkreft oppdaget ved screening 67% av all brystkreft blant inviterte kvinner i alderen 50-69 år.

**Mammografiundersøkelser utenfor Mammografiprogrammet**

Bruk av mammografi utenfor Mammografiprogrammet består av diagnostiske undersøkelser ved symptomer på brystkreft og mammografiundersøkelser av symptomfrie kvinner på kvinnens eget initiativ (opportunistisk screening). Omfanget av opportunistisk screening i Norge er ikke systematisk registrert. I følge Statens strålevern ble rundt 5 000 kvinner undersøkt med mammografi i 1983 og rundt 110 000 i 1993. En større andel kvinner har blitt undersøkt med mammografi i urbane strøk, særlig i Oslo. Omtrent 70% av undersøkelsene i 1993 ble utført ved private røntgeninstitutt. I spørreskjema som blir utdelt sammen med første invitasjon til Mammografiprogrammet, svarer over 60% at de har vært undersøkt med mammografi minst én gang tidligere. Det er ukjent hvor mye av dette som representerer diagnostisk mammografi og hvor mye som er opportunistisk screening. Det er ikke kjent hvor mange tilfeller av brystkreft som oppdages ved slik screening eller hvilken betydning dette har for dødeligheten av brystkreft.

**Evalueringsoppdrage**t

I 2006 ble det inngått avtale mellom HOD og Norges Forskningsråd (NFR) om at NFR skulle bidra til en forskningsbasert evaluering av det Mammografiprogrammet. NFR skulle utarbeide et mål- og rammedokument for evalueringen og gjennomføre evalueringen etter dette. Avtalen skulle vare ut 2010, men er forlenget til ut mai 2015 på grunn av forsinkelser av ulike årsaker. Det ble nedsatt en styringsgruppe (oversikt over medlemmene finnes i hovedrapporten) som skulle bistå NFR med utarbeidelse av Mål- og Rammedokumentet, utlysing av forskningsmidler til evalueringen og vurdering av søknader, vurdere de vitenskapelige arbeidene og utarbeide en sluttrapport til HOD.
Hovedmålene i Mål- og Rammedokumentet var:

- Evaluere effektiviteten av programmet på redusert dødelighet av brystkreft, eventuell endring i stadieinndeling av brystkreft, og i insidensen (forekomsten) av avansert brystkreft
- Evaluere organiseringen, tilgjengeligheten og kvaliteten av programmet, inklusive vitenskapelig aktivitet
- Økonomisk evaluering med kombinert bruk av ressurser og nytte/effektivitet av programmet

Prosjektmidler ble utlyst både nasjonalt og internasjonalt. Alle kunne søke med unntak av ansatte i det norske Kreftregisteret på grunn av mulige habilitetsproblemer. Elleve søknader, åtte med norske og tre med internasjonale prosjektledere (fra Sverige, England og Nederland), ble tatt i mot til medio april 2008. Et internasjonalt ekspert panel (deltakere fra Norge, Finnland, England og Nederland) bedømte den vitenskapelige kvaliteten på søknadene. På bakgrunn av disse vurderingene og relevansen til Mål- og Rammedokumentet, besluttet styringsgruppen at sju søknader kunne imøtekommes. Det ble lagt vekt på at minst to ulike forskningsgrupper skulle evaluere sentrale mål som reduksjon i dødelighet av brystkreft, overdiagnostikk og helseøkonomi. I tillegg dekket

søknadene intervallkreft og kvinners syn på og erfaringer med mammografiscreening. Ingen godkjente søknader inneholdt evaluering av endringer i stadieinndeling og organiseringen av Mammografiprogrammet.

For å kunne gjennomføre evalueringen måtte en ny prosjektdatabase etableres, basert på insidensdatabasen i Kreftregisteret og screeningsdatabasen i Mammografiprogrammet som var en pasientadministrativ database for invitasjoner, oppmøtte og resultat av screening, og eventuelt av etterinnkallinger og purring av invitasjoner til ikke møtte. Dette var et utfordrende arbeid som inkluderte blant annet variabelbeskrivelser og et omfattende kvalitetssikringsarbeid, og som medførte ett års utsettelse av evalueringsprosessen. Det tilkom også flere utsettelser som skyldtes manglende tilgang til informasjon om screenede kvinner med normalt mammografifunn og tillatelser til koblinger til andre registre som var nødvendig for å kunne gjennomføre evalueringen.

I tillegg til arbeidene og rapportene fra forskningsgruppene ble det gjort systematiske søk i PubMed/Medline etter originale, vitenskapelige arbeider, publisert i 2008 eller senere. Kravene var at det i tillegg skulle være fagfellevurderte publikasjoner, presentasjon av originale analyser og estimater basert på data fra Mammografiprogrammet eller andre originale data fra Norge. For publikasjoner av samme forfattere om samme tema, ble de arbeidene med lengst oppfølgingstid og som var av nyest dato, inkludert.

## Metodemessige utfordringer og vurderinger

Evalueringen har som mål å anslå nytten av å tilby mammografiscreening. Den samlede nytten vil avhenge både av effektiviteten til programmet og av hvor mange kvinner som møter til screening. Kvinner som møter til screening vil som gruppe kunne skille seg fra kvinner som ikke møter til screening blant annet når det gjelder faktorer som alder, utdanningsnivå og bruk av hormoner. Det er derfor grunn til å tro at kvinner som møter til screening ville hatt en annen risiko både for å utvikle brystkreft og for å dø av sykdommen, sammenliknet med kvinner som ikke møter, også dersom det ikke fantes et screeningtilbud. Slike forskjeller kalles ofte selv-seleksjon og kan føre til at observerte forskjeller mellom kvinner som møter og kvinner som ikke møter til screening, ikke nødvendigvis kan tilskrives effekten av programmet. Det er derfor å foretrekke å sammenlikne inviterte og ikke inviterte kvinner for å måle effekter av programmet.

### Effektivitet – redusert dødelighet av brystkreft

Mammografiscreening fremskynder diagnosetidspunktet for brystkreft med en ikke målbar periode som kalles ledetid (tiden fra svulsten blir oppdaget på screening til den ellers ville blitt oppdaget på grunn av symptomer). Dermed vil tiden fra diagnosetidspunktet til død av brystkreft øke for kvinner med svulster oppdaget på screening, uavhengig av om tidligere diagnose har ført til at kvinnen lever lenger med sykdommen. For å unngå denne feilkilden brukes vanligvis dødelighet av brystkreft i befolkningen for å vurdere effektiviteten av mammografiscreening. Det å bli invitert til mammografiscreening kan kun forebygge dødsfall av brystkreft som er oppdaget etter invitasjon. Derfor brukes ofte insidensbasert dødelighet av brystkreft (dødelighet av brystkreft som er oppdaget etter invitasjon) som mål på effektivitet.

## Overdiagnostikk

Påvisning av svulster ved screening som ikke ville blitt oppdaget i kvinnens levetid i en situasjon uten screening (overdiagnostikk) regnes som den største ulempen ved mammografiscreening. Siden det ikke er mulig å avgjøre om den enkelte svulst er overdiagnostisert eller ikke, vil overdiagnostikk føre til unødvendig behandling og psykisk belastning for den enkelte kvinne, og økt ressursbruk i helsevesenet.

Antall overdiagnostiserte vil være antall ekstra brystkrefttilfeller blant kvinner som tilbys screening sammenliknet med kvinner som ikke tilbys screening, når begge disse gruppene følges livet ut. For å beregne nivået av overdiagnostikk, er det viktig å ta hensyn til at en del av de svulstene som oppdages ved screening i alderen 50-69 år, ville blitt oppdaget fra 70 og oppover uten screening. Dette gjøres oftest ved å beregne nedgangen i forekomst av brystkreft fra 70 år og oppover blant kvinner som tidligere har vært invitert til screening, eller ved å gjøre antakelser om forventet gjennomsnittlig ledetid. Begge metoder forutsetter at forventet forekomst av brystkreft i fravær av screening kan beregnes til sammenlikning.

Det er ingen konsensus for hva ekstra antall brystkrefttilfeller bør sammenliknes med for å beregne andelen overdiagnostiserte. Det uavhengige britiske panelet UK Panel beskrev flere alternative nevnere som har vært brukt av forskere:

A Ekstra antall brystkreft som en andel av antall tilfeller diagnostisert i hele oppfølgingsperioden for kvinner som ikke inviteres til screening
B Ekstra antall brystkreft som en andel av antall tilfeller diagnostisert i hele oppfølgingsperioden for kvinner som inviteres til screening
C Ekstra antall brystkreft som en andel av antall tilfeller diagnostisert i screeningsperioden for kvinner som inviteres til screening
D Ekstra antall brystkreft som en andel av antall tilfeller diagnostisert ved screening

De anbefalte B for å beskrive samfunnsperspektivet (belastningen ved overdiagnostikk på befolkningsnivå) og C for å beskrive individperspektivet (risikoen for at brystkreft påvist mens kvinnen mottar screeninginvitasjoner, er overdiagnostisert). Beregninger av overdiagnostikk bør inkludere både infiltrerende brystkreft og DCIS fordi behandling av DCIS hos noen kvinner kan tenkes å forebygge senere infiltrerende brystkreft og hos andre kan tenkes å være overbehandling.

## Vurderinger av studiedesign og mulige kilder til systematiske feil

Den gradvise innføringen av Mammografiprogrammet tilsier at studier med individdata kan forventes å gi et mer pålitelig resultat enn studier med aggregerte (økologiske) data. Siden brystkreft i mange tilfeller er en sykdom som utvikler seg langsomt, vil lang oppfølgingstid være viktig for en pålitelig vurdering av både fordeler og ulemper. En særlig utfordring vil være å oppnå en balansert sammenlikning mellom inviterte og ikke inviterte kvinner på grunn av de forskjellene i forekomst og dødelighet av brystkreft som fantes mellom fylker som ble inkludert tidlig og sent i programmet. På samme måte vil de store endringene i bruk av hormoner og opportunistisk screening som foregikk samtidig med innføringen av Mammografiprogrammet kunne utgjøre vesentlige feilkilder i sammenlikninger over tid og mellom kvinner i ulike fylker.

# Resultater

## Dødelighet av brystkreft

Fem studier basert på original forskning og med uttalt mål om å vurdere om Mammografiprogrammet reduserer dødeligheten av brystkreft i Norge ble inkludert. Alle disse hadde insidensbasert dødelighet av brystkreft som utfall.

Studiene av Olsen og medarbeidere og Kalager og medarbeidere har store likheter i design med delvis økologiske data og sammenlikning av insidensbasert dødelighet av brystkreft etter innføringen av screening sammenliknet med insidensbasert dødelighet av brystkreft i historiske og regionale kontrollgrupper. Resultatene er også sammenliknbare med estimater på 7-11% reduksjon og 10% reduksjon i dødelighet av brystkreft for kvinner som ble invitert til screening. Kalager og medarbeidere beregnet absolutt risikoreduksjon til 2,4 dødsfall per 100 000 person-år. De største svakhetene i disse studiene er misklassifisering av eksponeringen ved bruk av økologiske data og manglende informasjon om opportunistisk screening, forskjeller mellom screeninggruppen og kontrollgruppene når det gjelder faktorer av betydning for dødeligheten av brystkreft, variasjon i antall screeningrunder for kvinner i den inviterte gruppen, kort oppfølgingsperiode og dermed lav statistisk styrke.

Weedon-Fekjær og medarbeidere utførte en åpen kohortstudie med individdata og sammenliknet dødeligheten av brystkreft blant inviterte kvinner og kvinner som ennå ikke hadde blitt invitert. De brukte metoder som ble utviklet spesifikt for å kunne studere insidensbasert dødelighet av brystkreft og samtidig utnytte mest mulig av de tilgjengelige data. Den beregnede reduksjonen i dødelighet av brystkreft for kvinner i alderen 50-79 år var 28%. De beregnet også at 368 kvinner måtte inviteres til screening gjennom 10 screeningrunder for å forebygge ett dødsfall av brystkreft blant kvinner 50-89 år. Den viktigste svakheten i denne studien er misklassifisering av eksponeringen på grunn av opportunistisk screening.

van Luijt og medarbeidere utførte en simuleringsstudie hvor individuelle livsløp med og uten screening ble simulert og sammenliknet med økologiske data på nasjonalt nivå for å oppnå best mulig tilpasning av modellen til norske forhold. I de to modellene som samlet sett ga best tilpasning for forekomst av brystkreft over tid, aldersspesifikk forekomst og relativ endring av dødelighet over tid, ble det beregnet at brystkreftdødeligheten for kvinner 55 - 80 år ville være redusert med 25 - 30% innen 2025. Disse beregningene var basert på screening i og utenfor Mammografiprogrammet sett under ett. Den absolutte effekten ble beregnet til ett forebygget dødsfall av brystkreft per 1 470 - 2 612 utførte screeningundersøkelser. Den største usikkerheten ved disse resultatene er knyttet til de antakelsene om risikofaktorer for brystkreft blant norske kvinner og om tumorvekst som var nødvendige å inkludere i modellen for å oppnå en god tilpasning.

Hofvind og medarbeider sammenliknet dødelighet av brystkreft blant inviterte kvinner som møtte og inviterte kvinner som ikke møtte til undersøkelse i Mammografiprogrammet. All informasjon var målt på individnivå. De beregnet at kvinner som møtte hadde 43% lavere dødelighet av brystkreft enn kvinner som ikke møtte og at denne forskjellen tilsvarte 36% lavere dødelighet for inviterte kvinner. Selv om forskerne forsøkte å korrigere for selv-seleksjon til screening, utgjør dette likevel en stor svakhet i studien. En annen svakhet er at det ikke kunne tas hensyn opportunistisk screening.

Studier tyder på at mammografi har vært hyppig brukt utenfor programmet, men det er ikke mulig å si med sikkerhet hvor hyppig eller hvor mye av

mammografivirksomheten utenfor programmet som kan betraktes som erstatning for programmet. Hvor mye misklassifisering den opportunistiske screeningen representerer er derfor usikkert.

*Konklusjoner*
I studiene av insidensbasert dødelighet av brystkreft ble det funnet en større reduksjon i brystkreftdødelighet i studier med detaljert informasjonen om invitasjonsdato og lang oppfølgingstid sammenliknet med studier hvor bostedsfylke eller invitasjonskrets ble brukt som en tilnærming til invitasjonsstatus og hvor oppfølgingsperioden var kort. Manglende informasjon om omfanget og betydningen av opportunistisk screening er en svakhet ved alle studiene. I noen av studiene vil forskjeller i forekomst og dødelighet av brystkreft over tid og mellom ulike deler av landet også bidra til usikkerheten.

Den beregnede reduksjonen i dødelighet av brystkreft var 7%, 10%, 28% og 30% i de fire studiene med mest pålitelig design og metoder. De to studiene med lavest estimat har delvis økologisk design med stor risiko for misklassifisering og kort oppfølging. Av de to gjenværende studiene er simuleringsstudien basert på antakelser om risikofaktorer og tumorvekst som er vanskelige å sannsynliggjøre, mens studien til Weedon-Fekjær og medarbeidere inkluderer mest mulig av den tilgjengelige informasjonen gjennom individuell invitasjonsdato og lang oppfølgingstid. Betydningen av ulike antakelser for resultatet ble også testet. Denne studien ble vurdert av styringsgruppen som den mest pålitelige. Vår vurdering er derfor at det mest pålitelige estimatet for reduksjon i dødelighet av brystkreft som kan tilskrives innføringen av Mammografiprogrammet er mellom 20 og 30% for kvinner i alderen 50-79 år. Denne reduksjonen gjelder for en situasjon med et screeningprogram for kvinner 50-69 år med oppfølging til 79 år sammenliknet med en situasjon uten et screeningprogram. Estimatene tyder på at Mammografiprogrammet som helhet fungerer slik det kunne forventes basert på de fleste tidligere systematiske gjennomganger av randomiserte kontrollerte forsøk.

**Overdiagnostikk**

Åtte publikasjoner, hvorav fem fra forskergruppene i evalueringen, fra 2008 og fremover som presenterte estimater på overdiagnostikk av brystkreft etter introduksjonen av Mammografiprogrammet ble inkludert.

Det er stor variasjon i den rapporterte andelen overdiagnostiserte mellom de ulike studiene, noe som dels kan forklares av ulike nevnere (beskrevet over). En annen årsak til de ulike estimatene er at noen studier har inkludert DCIS, mens andre kun har studert infiltrerende brystkreft.

Tre studier med individdata sammenliknet brystkreftrisikoen blant møtte og ikke-møtte kvinner og inkluderte både infiltrerende brystkreft og DCIS. To studier av Lund og medarbeidere hvor det ble brukt data fra undersøkelsen Kvinner og kreft ga nokså ulike resultater for kvinner som møtte i Mammografiprogrammet sammenliknet med kvinner som aldri hadde vært til mammografi (18% og 7% av kvinner i alderen 50-79 år *med* screening – B). Forskjellen kan skyldes manglende presisjon, selvrapportert oppmøte i en av studiene, forskjeller i selv-seleksjon og justering for risikofaktorer. I den minste av de to studiene hvor det ble funnet 18% kunne heller ikke prevalensscreeningen inkluderes.

Sørum Falk og medarbeidere brukte landsdekkende data på individnivå til å beregne andelen overdiagnostiserte i en tenkt gruppe kvinner som møtte regelmessig til mammografi fra 50 til 69 år og ble fulgt livet ut. Oppmøte ble beregnet å gi en livstidsrisiko for overdiagnose på mellom 16,5% og 19,6%, omregnet til 13,9-16,5% for

inviterte kvinner sammenliknet med kvinner som ikke var screenet (A). Muligheten for selv-seleksjon kan ikke utelukkes for noen av disse tre studiene.

Duffy & Michalopoulos brukte en kombinasjon av observerte data og modellering av ledetid gjennom to litt ulike tilnærminger, og fant at 15-17% av infiltrerendebrystkreft og DCIS oppdaget på screening kunne tilskrives overdiagnostikk (D). Estimering av ledetid basert på insidensen av brystkreft i det første året etter screening i kombinasjon med at det ikke ble tatt hensyn til konkurrerende risiko (død av andre årsaker) har trolig ført til underestimering av andelen overdiagnostiserte.

van Luijt og medarbeidere utførte en simuleringsstudie med tilpasning til norske forhold ved hjelp av norske aggregerte data. De beregnet overdiagnostikk samlet for infiltrerende brystkreft og DCIS til henholdsvis 2-11% og 3-19% av brystkreft i en situasjon med screening for kvinner i alderen 50-100 år (B) og 50-70 år (C) i perioden 2000-2009 og 2-7% og 3-11% i perioden 2014-2023 for samme aldersgrupper. De høyeste estimatene ble funnet i modeller som antok lengst ledetid. Den største usikkerheten i studien er knyttet til de forutsetningene som ble gjort for å oppnå en tilfredsstillende tilpasning av modellen til norske data.

Kalager og medarbeidere brukte et delvis økologisk design med sammenlikning av ulike fylker over tid for å studere overdiagnostikk av infiltrerende brystkreft. I den analysen som hadde lengst oppfølgingstid fant de 18% overdiagnostikk blant kvinner 50-79 år sammenliknet med en situasjon uten screening – nevner A). Selv om studien i noen grad kunne ta hensyn til økningen i insidens i utenom screening, kunne den ikke ta hensyn til forskjeller i utvikling over tid mellom ulike fylker eller misklassifisering som følge av opportunistisk screening. Studien kunne også bare delvis ta hensyn til den forventede lavere forekomsten av brystkreft blant kvinner i alderen 70-79 år på grunn av kort oppfølgingstid og misklassifisering av invitasjonsstatus blant kvinner i denne aldersgruppen.

Zahl & Mæhlen og Jørgensen & Gøtzsche beregnet overdiagnostikk i økologiske trendstudier. Overdiagnostikk ble beregnet som en andel av brystkreft blant kvinner i alderen 50-69 år i en situasjon *uten* screening, det vil si en femte nevner sammenliknet med de fire som ble beskrevet av UK Panel. Siden denne nevneren vil være lavere enn i en situasjon med screening, vil andelen overdiagnostiserte nødvendigvis bli høyere enn C selv om telleren beregnes utfra den samme informasjonen. Zahl & Mæhlen rapporterte 50% overdiagnostikk av infiltrerende brystkreft, mens Jørgensen & Gøtzsche rapporterte 37% for infiltrerende brystkreft og 52% for infiltrerende brystkreft og DCIS samlet. Delvis inklusjon av den forventede lavere forekomsten av brystkreft etter screeningalder kan ha bidratt til et lavere estimat hos Jørgensen & Gøtzsche sammenliknet med Zahl & Mæhlen. Begge estimatene er trolig overestimert på grunn av misklassifisering av invitasjonsstatus blant kvinner i alderen 70-79 år, samt manglende justering for faktorer som hormonbehandling og opportunistisk screening i beregning av insidensraten uten screening.

Suhrke & Zahl undersøkte risikoen for brystkreft blant kvinner som fikk hormonbehandling mot plager i overgangsalderen og fant en dobling i risiko for kvinner som brukte slike medikamenter over lengre tid sammenliknet med kvinner som ikke fikk hormonbehandling.

*Konklusjoner*

Det beregnede nivået av overdiagnostikk var lavere i studier med individuell informasjon om invitasjonsdato og i studier hvor insidensfallet etter screening ble estimert blant kvinner som alle hadde vært invitert til screening da de var yngre. Studier som justerte for

ledetid som et alternativ til å inkludere insidensfallet etter screening fant også lavere nivå av overdiagnostikk i disse analysene. Betydningen av opportunistisk screening og hormonbehandling er en kilde til usikkerhet i alle studiene og kan ha bidratt til utfordringene med å estimere insidensrater i fravær av screening.

Variasjonen i design, analytisk tilnærming og valg av nevnere i studiene, både i og utenfor evalueringens prosjekter, kan forklare det meste av variasjonen i estimatene fra 7% til 52%, og gjør det vanskelig å anslå et samlet estimat. De økologiske og delvis økologiske studiene er beheftet med misklassifisering av invitasjonsstatus og noen av dem har også kort oppfølgingstid. Studier med bruk av individdata kombinert med lengre oppfølging og inklusjon av både infiltrerende brystkreft og DCIS, forventer vi vil gi de mest pålitelige resultatene.

Vår vurdering er at de mest pålitelige estimatene for overdiagnostikk av infiltrerende brystkreft og DCIS samlet for kvinner i alderen 50-79 år sammenliknet med en situasjon *uten* screening (A), er i størrelsesorden 15-25%. Blant kvinner i alderen 50-79 år i en situasjon *med* screening tilsvarer dette 15-20% (B). De øvrige nevnerne som ble brukt i studiene (C, D, Zahl & Mæhlen og Jørgensen & Gøtzsche) inkluderer kun brystkreft i screeningalderen og ikke perioden med insidensfall etter screening. Basert på informasjon i de ulike studiene anslår vi at andelen overdiagnostiserte blant kvinner som inviteres til screening (C) er i størrelsesorden 20-30%.

**Intervallkreft**

De studiene som har inkludert landsdekkende data fra flest mulig screeningrunder rapporterer intervallkreftraten for infiltrerende brystkreft og DCIS samlet til 1,7/1 000 undersøkelser, hvorav rundt 30% oppdages i det første året etter screening. Intervallkreft utgjorde rundt 25% av krefttilfellene blant kvinner som møtte til screening.

Tre studier hvor andelen falske negative blant intervallsvulstene ble undersøkt, tyder på at 30-35% av intervallkrefttilfellene kunne vært oppdaget på forrige screening, mens resten ikke kunne påvises på tidligere mammografier. Til sammenlikning kunne også 20% av de screeningoppdagede svulstene gjenfinnes på foregående mammografi.

Intervallsvulstene var større og var oftere av lobulær type enn svulster oppdaget blant kvinner som ennå ikke hadde blitt invitert til screening. Dette gjaldt særlig de svulstene som hadde blitt oversett ved forrige screeningundersøkelse. Prognosen for kvinner med intervallkreft var lik prognosen for kvinner som fikk påvist brystkreft før de ble invitert til screening første gang. I disse studiene kan opportunistisk screening og selv-seleksjon ha hatt betydning for resultatet.

Intervjuer med kvinner som hadde fått påvist intervallkreft tyder på at noen kvinner venter med å søke lege når de utvikler tegn på brystkreft fordi de har hatt en normal mammografi i forrige screeningrunde eller venter på neste screeningrunde. Kvinnene som ble intervjuet var fortsatt positivt innstilt til Mammografiprogrammet selv om de hadde fått påvist kreft mellom screeningrundene.

**Etterundersøkelser, falske positive undersøkelser og erfaringer med deltakelse i Mammografiprogrammet**

Roman og medarbeidere estimerte at 20% av kvinner som møter til 10 screening-runder fra 50 til 69 år kan forventes å bli innkalt til etterundersøkelser minst én gang, med konklusjon om at det ikke kan påvises brystkreft og at mammografiscreeningen dermed var falsk positiv. 4% vil bli undersøkt med invasive tester (biopsi eller cytologi) før brystkreft kan utelukkes. Siden informasjon fra 10 screeningrunder ennå ikke er

tilgjengelig, ble beregningene basert på en antakelse om at andelen etterundersøkte var den samme i de siste fire screeningrundene. Det var stor variasjon i andelen etterundersøkte mellom de ulike screeningsentrene, noe som kan skyldes forskjeller i underliggende brystkreftforekomst, antall screeningrunder som har vært gjennomført ved hvert senter, men også forskjeller i praksis og kvalitet på undersøkelsene.

I to studier ble den psykiske påkjenningen ved etterundersøkelser kartlagt. Disse studiene tyder på at den psykiske belastningen er høy i perioden rundt selve undersøkelsen, men at påkjenningen avtar over tid. Kvinner som fikk påvist brystkreft på etterundersøkelsen hadde høyere depresjonsscore i perioden etter undersøkelsen. Antall deltakere i disse studiene var lavt og frafallet gjennom oppfølgingsperioden var stort, noe som øker risikoen for seleksjonsskjevhet. Intervjuer med kvinner som ble innkalt til etterundersøkelser støtter funnene om at en etterundersøkelse er en psykisk belastning og tyder også på at kvinnene ikke er innstilt på at de kan bli innkalt til etterundersøkelser. Informasjonen som ble gitt sammen med innkallingen virket beroligende på noen og økte bekymringen hos andre. De fleste kvinnene var fortsatt positivt innstilt til screening etter innkallingen.

Gruppeintervjuer med screeningdeltakere viser at kvinner er klar over at svulster kan bli oversett ved mammografi, men at de likevel stoler mer på screeningen enn på selvundersøkelse eller klinisk undersøkelse hos lege. Invitasjonen opplevdes mer som en innkalling enn en invitasjon og for de intervjuede kvinnene var det derfor lite behov for en beslutningsprosess. De fleste kunne ikke huske at de hadde mottatt informasjon om screeningen, men mente likevel at de visste nok til å ta beslutningen. Hensikten med å delta var primært å få bekreftelse på at de ikke hadde brystkreft. Deltakelse gjennom flere screeningrunder økte opplevelsen av rutine og av å være en del av et «samlebånd». Smerte og psykisk belastning ble mindre i senere screeningrunder sammenliknet med første undersøkelse.

### Kostnader og kostnad-effekt-beregninger

Moger & Sønbø Kristiansen beregnet kostnader per screeningrunde til 574 millioner kroner, eller 1 389 kroner per undersøkte kvinne i 2012, inkludert kostnader ved screeningundersøkelser og etterundersøkelser, samt indirekte kostnader. Øvrige kostnader knyttet til drift av Mammografiprogrammet ble ikke inkludert, og de faktiske kostnadene knyttet til programmet er derfor høyere. Moger og medarbeidere estimerte kostnader ved behandling av brystkreft i 2008 basert på den behandlingen som ble gitt i 2008-2009. Kostnader ved 10 års behandling av en kvinne med brystkreft ble beregnet til rundt 356 000 kroner. Kostnadene var lavere for brystkreft oppdaget på screening og høyere ved intervallkreft og kreft blant kvinner som ikke deltok eller ikke hadde blitt invitert til screening. Kostnadene var også avhengig av sykdommens utbredelse ved diagnosetidspunktet, med høyere kostnader for mer avansert sykdom.

Resultatene fra studiene beskrevet over ble brukt i en simuleringsstudie for å beregne kost-nytte-balanse i form av kostnader per kvalitetsjusterte leveår. Beregninger av kostnad-effekt-brøk avhenger av både kostnader og effektivitet, men også av andre faktorer som oppmøte, sensitivitet av mammografien, forekomst av brystkreft i befolkningen, nivå av overdiagnostikk og overbehandling, omfanget av opportunistisk screening og behandling av brystkreft. I tillegg avhenger antall kvalitetsjusterte leveår av nettogevinsten ved redusert sykelighet og dødelighet og hvilken livskvalitet som blir tillagt ulike sykdomsfaser og varigheten av disse. van Luijt og medarbeidere beregnet kostnader per kvalitetsjusterte sparte leveår ved screening til mellom 190 000 og 479 000

norske kroner. I andre sammenhenger har Helsedirektoratet brukt 400 000 – 1 million norske kroner som kostnadsterskel per sparte leveår.

### Fordeler og ulemper

Med utgangspunkt i resultatene beskrevet overfor og informasjon fra de omtalte studiene, har vi forsøkt å balansere forventede fordeler og ulemper ved screening for 10 000 kvinner som mottar sin første invitasjon til screening 50 år gamle og inviteres til totalt 10 screeningrunder frem til 69 år. Vi antar at 76%, det vil si 7 600 kvinner, møter regelmessig, mens de resterende 2 400 ikke møter i det hele tatt. Blant de som møter vil rundt 377 kvinner få påvist infiltrerende brystkreft eller DCIS på screening. Av disse vil rundt 27 kvinner unngå å dø av brystkreft som følge av tidlig diagnose og behandling, mens rundt 142 kvinner forventes å få påvist sykdom som ikke ville blitt oppdaget i kvinnens levetid uten screening (overdiagnostikk). Dette tilsvarer i overkant av 5 overdiagnostiserte for hver kvinne som reddes fra å dø av brystkreft. Som gruppe vil overdiagnostiserte svulster forventes å være mildere enn annen brystkreft og kreve mindre aggressiv behandling.

Rundt 127 kvinner forventes å få påvist brystkreft mellom to screeningundersøkelser – intervallkreft. I ettertid ville omtrent 42 av disse svulstene kunne påvises på forrige mammografibilde og dermed regnes som oversette (falske negative). Omtrent 1 520 kvinner forventes å bli innkalt til etterundersøkelser minst én gang uten at det blir påvist brystkreft (falske positive). For de fleste av disse vil en ny mammografi eller ultralyd være tilstrekkelig til å utelukke kreft, mens for rundt 310 kvinner vil invasive tester (biopsi eller cytologi) være nødvendig. De øvrige deltakerne (omtrent 5 576 kvinner) vil ha sanne negative undersøkelser gjennom alle 10 screeningundersøkelsene.

### Gjenværende usikkerhet

Siden denne evalueringen er basert på observasjonsstudier, har metodemessige vurderinger spilt en sentral rolle i oppsummeringen av resultater. Selv i de studiene som ble vurdert som mest pålitelige, vil det være flere sentrale faktorer som ikke kunne tas hensyn til.

Det er klart at opportunistisk screening har foregått i et tilstrekkelig omfang til å utgjøre en potensielt viktig feilkilde i de fleste studiene i denne evalueringen. Særlig har det å beregne forventet forekomst av brystkreft i fravær av screening vist seg å være utfordrende, også på grunn av de store endringene i bruk av hormonbehandling som har foregått i samme periode. Informasjon om hormonbruk på individnivå har kun vært tilgjengelig for den siste delen av studieperioden.

Omorganisering og sentralisering av brystkreftbehandling foregikk parallelt med innføringen av Mammografiprogrammet. I de fleste studiene hvor dødelighet av brystkreft ble undersøkt, var det vanskelig å skille betydningen av disse endringene fra betydning av selve programmet.

Flere forskere beskrev problemer å tilpasse prediksjonsmodeller til den observerte forekomsten av brystkreft, mens andre brukte enkle metoder uten noen vurdering av tilpasningsgraden. Usikkerheten rundt forekomsten av brystkreft i fravær av screening kompliserer tolkningen av de beregnede nivåene av overdiagnostikk. Studier av møtte og ikke-møtte kvinner vil i mindre grad være avhengig av trendberegninger. Disse studiene vil derimot være sårbare for påvirkning av selv-seleksjon til screening. Dette medfører at

ingen av de enkelte studiene om overdiagnostikk kunne sies å gi det mest pålitelige resultatet, og konklusjonene om nivået av overdiagnostikk må derfor betraktes som svært usikre.

En lang oppfølgingsperiode er nødvendig for å kunne måle både fordeler og ulemper ved mammografiscreening. Den lengste oppfølgingen av kvinner som hadde blitt invitert til Mammografiprogrammet i studiene som ble inkludert i denne evalueringen var 14 år. I de fleste fylkene var oppfølgingen mye kortere. Per i dag har ingen årskull blitt invitert til 10 screeningrunder. De årskullene som det har vært mulig å følge i en periode etter at de har gått ut av screeningprogrammet på grunn av alder, har kun rukket å bli invitert til noen få screeningrunder. Dette innebærer at ingen av studiene i evalueringen har kunnet følge kvinner gjennom hele screeningprogrammet og livet ut. Både beregnede fordeler og ulemper kan forandre seg med lengre oppfølging.

Beregninger av kostnad-effekt-brøk avhenger av pålitelig informasjon om både fordeler, ulemper og kostnader. I tolkningen av kostnad-effekt-brøken er det viktig å merke seg at det nivået av overdiagnostikk som ble lagt til grunn i kostnad-effekt-analysene var lavere enn det nivået av overdiagnostikk styringsgruppen har anslått basert på alle studiene i evalueringen. Et høyere nivå av overdiagnostikk forventes å gi en større kostnad-effekt-brøk.

Studiene av kvinners erfaringer med screening og etterundersøkelser inkluderte kun kvinner som møtte til screening. Resultatene fra disse studiene bør derfor ikke betraktes som representative for alle inviterte kvinner.

## Anbefalinger og konklusjon

Resultatene fra denne evalueringen tyder på at Mammografiprogrammet som helhet fungerer slik det kunne forventes basert på de fleste tidligere systematiske gjennomganger av randomiserte kontrollerte forsøk. Tross stor usikkerhet rundt resultatene, ser det fra et samfunnsperspektiv ut til at balansen mellom kostnader og effekt i Mammografiprogrammet er innenfor det nivået helsemyndighetene definerer som akseptabelt for helsetjenester. For den enkelte kvinne som skal bestemme seg for om hun vil delta eller ikke, er det viktig å vurdere fordeler og ulemper utfra verdier, helse og livssituasjon.

Den store usikkerheten i denne evalueringen skyldes blant annet manglende kunnskap om mammografiscreening utenfor programmet og en kort oppfølgingsperiode etter invitasjon. Videre kan verken fordeler eller ulemper betraktes som konstante. Vi anbefaler derfor at det utarbeides en plan for fortsatt evaluering av Mammografiprogrammet. Arbeidet med kvalitetssikring av programmets database bør videreføres for å sikre mulighet for slik evaluering. Vi anbefaler også at kunnskapen om de utfordringene ved evalueringen av Mammografiprogrammet som er beskrevet i rapporten brukes i evalueringsplaner ved innføring av nye helsetjenestetilbud.

# 1 Introduction

This introduction contains a short description of breast cancer occurrence and treatment in Norway, an overview of general screening principles and of the Norwegian Breast Cancer Screening Program. The aim is to provide the non-expert with a background to the evaluation.

## 1.1 Breast cancer in Norway

### 1.1.1 Invasive breast cancer

Breast cancer is the most common form of cancer among women in Norway [1]. In 2012, there were 2956 new cases of invasive breast cancers for women of all ages combined [2]. Based on the incidence[1] in 2012, it is estimated that the risk of getting a breast cancer diagnosis before age 85 years is 10.5% for Norwegian women[2]. The incidence rates[3] of breast cancer for women in different age groups have changed considerably over the last decades. In the 1980s and early 1990s, breast cancer incidence rates increased steeply with increasing age until 45-50 years (figure 1a), followed by a constant or even modestly declining rate and finally a more modest increase from age 55-60 years. In the later years, this pattern has changed, and the steep increase now continues up to a higher age, and is followed by a drop in rates around 70 years, before the rates continue to increase. The same changes are also reflected in trends in incidence rates for women in defined age groups (figure 1b). Until the 1980s, incidence rates have increased steadily in all age groups except the youngest. From the mid-1990s there was a rapid increase in incidence rates for women 50-69 years, followed by a decline from the early 2000s.

Breast cancers are very heterogeneous in terms of their clinical, morphological and molecular profiles, and can be grouped according to a number of prognostic factors. The extent of disease at the time of diagnosis is classified according to tumor stage, which involves evaluation of time-dependent characteristics such as tumor size and involvement of adjacent tissue, lymph node invasion and distant metastasis [3]. Morphologically, breast cancer can also be classified according to histological type, which reflects tumor

[1] *Breast cancer incidence* is the number of new breast cancers during a specified time period in a defined population. Example: The number of invasive breast cancers among Norwegian women aged 50-69 years in 2012 was 1524.

[2] *Breast cancer risk* is the number of new breast cancers in a population during a specified time period divided by the total number of individuals in the population at the start of the time period. Also termed *breast cancer incidence proportion*.

[3] *Breast cancer incidence rate* is the number of new breast cancers in a population during a specified time period divided by the total amount of person-time at risk for of developing breast cancer in the same population during the same time period. Example: The incidence rate of breast cancer for Norwegian women aged 50-69 years in 2012 was 263/100 000 person-years (1524 breast cancer cases / 580 102 person-years).
*Person-time* is the time at risk of developing a disease (or another event of interest) for a person followed over time. The total person-time in the denominator of a rate should be the sum of the person-time for each person in the population during the specified time period. Example: 1 person followed for 50 years and 2 persons followed for 25 years each both result in 50 person-years.

growth patterns and nuclear grade, which incorporates tumor differentiation and proliferation. Overall, 60-75% of invasive cancers are ductal carcinomas of no special type, whereas lobular carcinomas constitute 5-15%, and the remaining types, such as tubular, medullary, mucinous

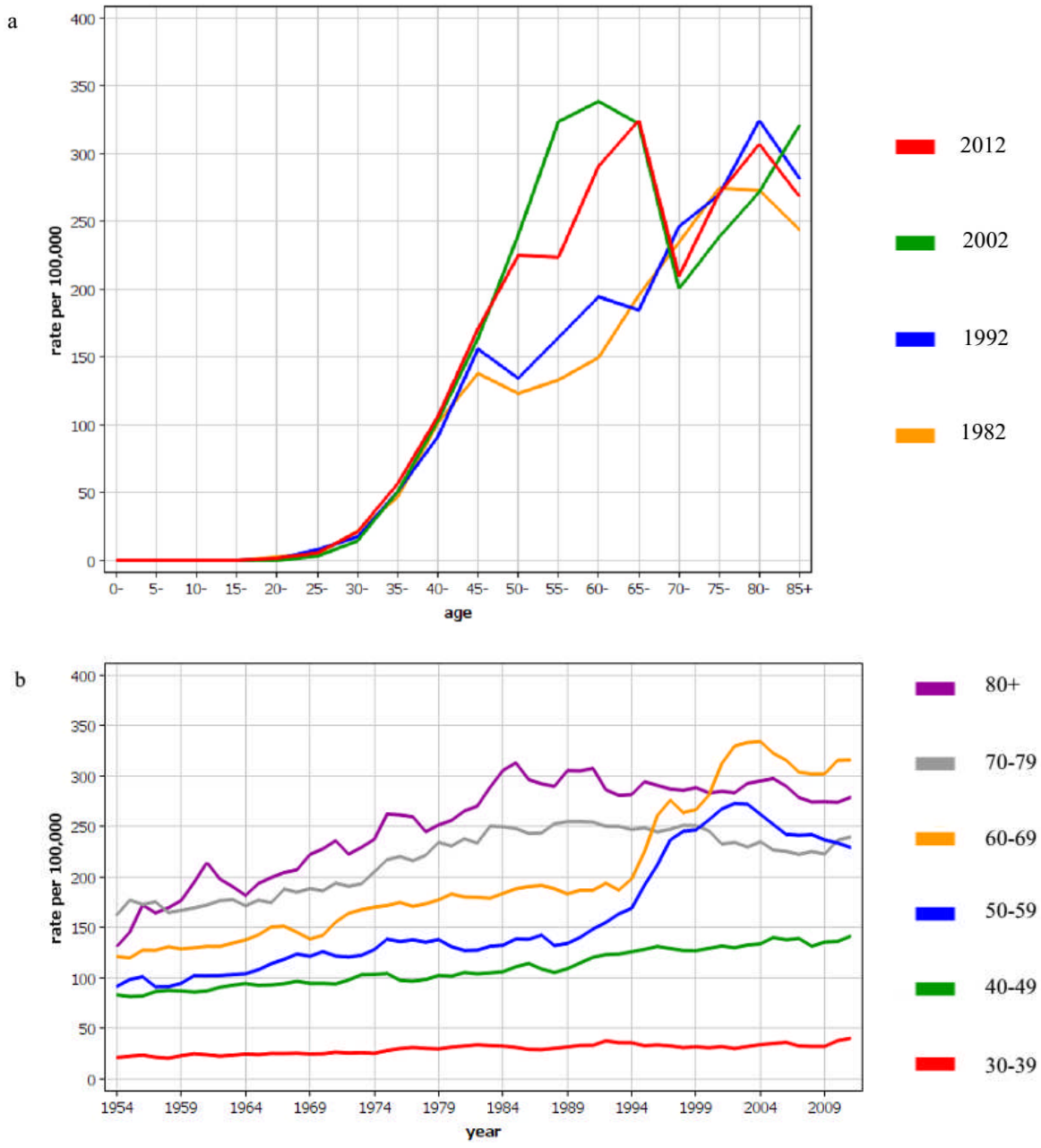



Figure 1. Incidence rates of invasive breast cancer in Norway by age (a) and time period (b).

Engholm G, Ferlay J, Christensen N, Kejs AMT, Johannesen TB, Khan S, Milter MC, Ólafsdóttir E, Petersen T, Pukkala E, Stenz F, Storm HH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 7.0 (17.12.2014). Association of the Nordic Cancer Registries. Danish Cancer Society. Available from http://www.ancr.nu, accessed on 09/01/15.

and papillary carcinomas, are rarer and comprise 1-5% each [4, 5]. In assessment of nuclear grade, tumors are scored according to the degree of tubule formation, nuclear polymorphism and mitotic activity and categorized into grade 1-3 from least to most aggressive [6]. Molecular classification of breast cancer is rapidly expanding, and over the last decades a range of protein and gene expression studies have revealed new prognostic markers, underscoring the heterogeneity of breast tumors [7-9]. However, only a limited number of these markers have so far been included in treatment algorithms, namely hormone receptor status, ERBB2/HER2 oncogene amplification and Ki-67 expression [10, 11].

### 1.1.2 Ductal carcinoma in situ

Ductal carcinoma in situ (DCIS) is a pre-invasive neoplastic lesion in the breast and is considered a non-obligate precursor of invasive breast cancer [12]. In the prescreening period, DCIS represented about 5% of all malignant breast tumors, compared to 10-20% in populations offered mammography screening [13, 14]. Similar to invasive breast cancer, DCIS is a heterogeneous condition, displaying many of the morphologically and molecular features found in invasive cancers [15], and can be classified according to nuclear grade [16]. The proportion of DCIS that eventually progresses to invasive cancer is difficult to study since many in situ tumors can go undetected [17], and since surgical removal is obligatory for those that are detected [18]. Small case series of women with symptomatically detected but untreated DCIS indicate that around 40% of these women will develop invasive breast cancer in the same area of the breast within 30 years [19, 20]. When treated, development of invasive cancer 10-15 years after surgical removal of DCIS has been shown to occur in 2-20% of the women, depending on type of treatment and tumor characteristics [18, 21]. Risk of death from breast cancer is also higher than in the general population, with estimates of 2-2.5% of patients dying from breast cancer within 10 years from diagnosis of DCIS [18, 21].

### 1.1.3 Breast cancer treatment and management

Over the last four decades, important improvements in treatment and management of breast cancer have occurred. This chapter will focus mainly on changes in treatment guidelines in Norway, as a basis for understanding the contribution of improved treatment to changes in mortality over the same period.

When the first national guidelines for treatment of breast cancer in Norway were established in 1981, modified radical mastectomy including removal of axillary lymph nodes was the recommended surgical treatment [22]. As breast-conserving surgery in combination with radiation therapy became more frequent, recommendations for when breast conserving surgery could be recommended without expected reduction in survival, were developed. Sentinel lymph node biopsy became part of clinical practice in 2000-2001 [23, 24]. Changes in surgical treatment have mainly been towards less extensive procedures, and would primarily be expected to increase quality of life without reducing survival expectations.

Adjuvant hormonal treatment with the anti-estrogen tamoxifen was initially recommended only for advanced stage disease (T3-T4 or N2-N3) among women with estrogen receptor positive tumors, with treatment duration of 3 years [22]. Indications were changed over the following decade, to include all node positive cases from 1988

[25], restriction to nuclear grade 2-3 from 1994 [26], and inclusion of T1c tumors from 2003 [23]. Treatment duration was 2 years from 1988 and extended to 5 years from 1998 [25, 27]. Current guidelines recommend hormonal treatment to all patients with hormone receptor positive tumors, except for postmenopausal patients with grade 1, small tumors (G1, pT1a-b) [11]. Aromatase inhibitors were introduced in adjuvant therapy in 2005.

Changes in adjuvant cytostatic chemotherapy regimens have been substantial. The first recommendations consisted of cyclophosphamide, vincristine, and methotrexate or 5-fluorouracil, given the day of surgery and one week after [22]. During the 1990s, recommendations were expanded to 6 months treatment with cyclophosphamide, methotrexate and 5-fluorouracil for patients with localized breast cancer, initially for patients younger than 55 years, with extension to patients younger than 65 years in 1998 [27]. From 2000, anthracycline replaced methotrexate for younger patients [24], and from 2003 for all patients offered chemotherapy [23]. This combination is still the basis for chemotherapy regimens in current Norwegian guidelines, but with the addition of trastuzumab (since 2005) and taxane (since 2008) according to HER2 and Ki67 expression, and removal of the previous upper age limit for chemotherapy [11].

Radiation therapy was initially recommended to patients with four or more affected lymph nodes and when complete surgical removal of the tumor was not obtained [22]. Since 1999, involvement of just one lymph node has led to radiotherapy [24]. From the implementation of breast-conserving surgery, radiation therapy towards the remaining breast tissue has been recommended to all women undergoing this treatment. The main indications have remained largely unchanged, but with differentiation of the radiation techniques and regimens over time.

For non-curable stages of breast cancer, the first treatment algorithms for endocrine treatment and chemotherapy were introduced in 2002 and 2004, respectively. There are also treatment algorithms according to HER2 status [11].

Treatment of DCIS has been either mastectomy or wide excision/breast-conserving therapy followed by radiation therapy, depending on tumor size and growth pattern [11].

### 1.1.4 Death from breast cancer

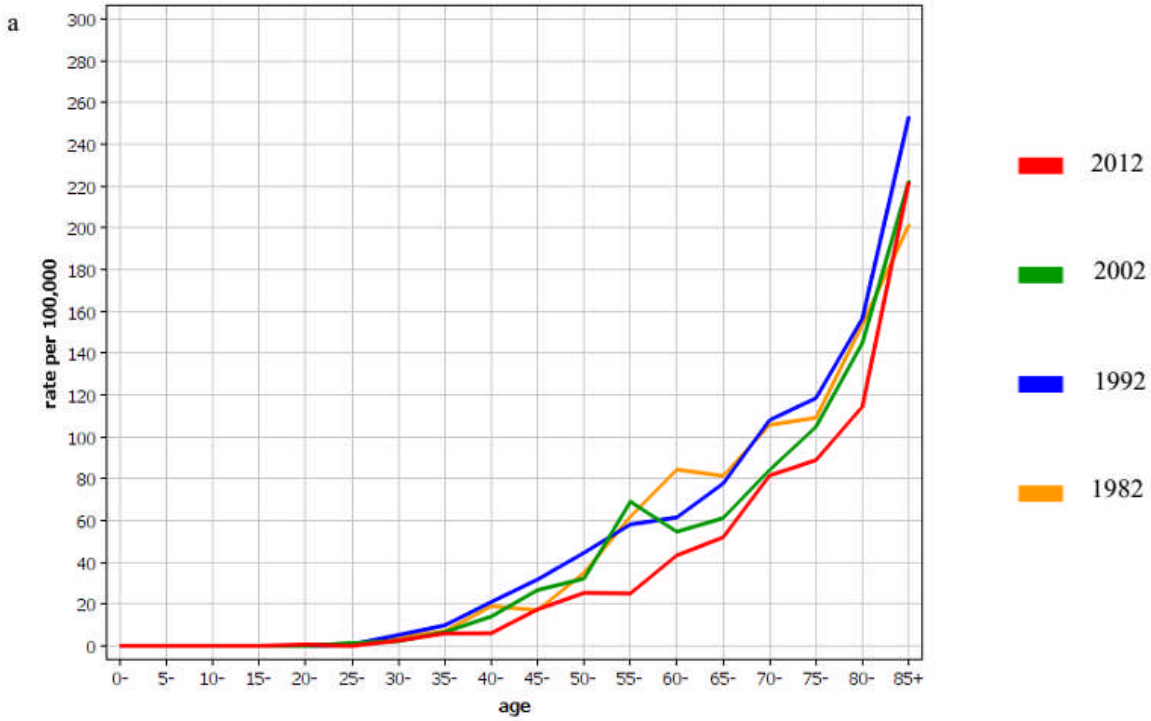
The mortality rates[4] from breast cancer increases steeply with age (Figure 2a). Until the mid-1990s, the mortality rates from breast cancer were stable in all age groups up to 69 years (Figure 2b). Among older women, breast cancer mortality rates increased until the mid-1990s. Since then, there has been a decline in breast cancer mortality across all age groups.

In the early 1990s, there were approximately 660 deaths from breast cancer per year among women 50 years and older in Norway, compared to 588 deaths from breast cancer in 2012. The total number of deaths among women in this age group was 21 178

[4] *Breast cancer mortality rate* is the number of deaths from breast cancers in a population during a specified time period divided by the total amount of person-time in the same population during the same time period. Unless otherwise specified, the number of deaths includes all deaths occurring in the population during the specified time period, regardless of when the cancer had been detected. Example: Breast cancer mortality among Norwegian women aged 50-69 years in 2012 was 35.5/100 000 person-years. This includes all deaths in women 50-69 years due to breast cancer in 2012 in Norway, regardless of whether the cancer was detected in 2012 or earlier and whether the cancer was detected while the woman was aged 50-69 or younger.

[28], and deaths from breast cancer thus constituted 2.6% of all deaths among women 50 years or older in 2012.

Survival after a diagnosis of breast cancer has increased over several decades. In the 1960s, women aged 50-69 years at diagnosis had a 10-year relative survival[5] of approximately 50% (all stages combined), increasing to more than 80% in the 2000s. Women who died from breast cancer in 1991-1995, had a median time from diagnosis to death of 5.5 years [29].




[5] *Relative survival from breast cancer* is the observed proportion of breast cancer patients still alive at a specified time after diagnosis divided by the expected proportion alive after the same amount of time in a comparable group in the general population.

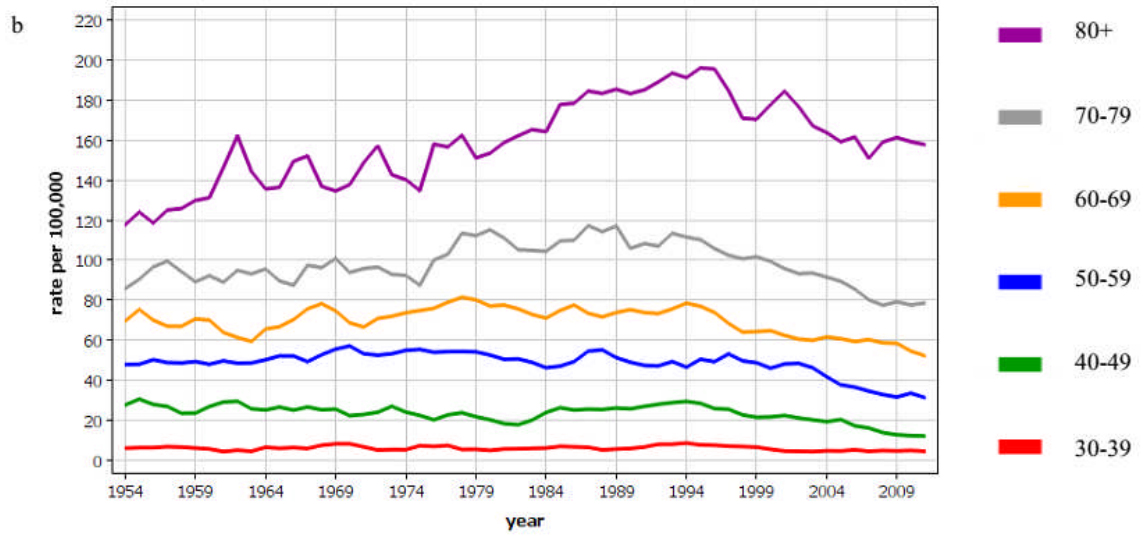



Figure 2. Breast cancer mortality rates in Norway by age (a) and time period (b).

Engholm G, Ferlay J, Christensen N, Kejs AMT, Johannesen TB, Khan S, Milter MC, Ólafsdóttir E, Petersen T, Pukkala E, Stenz F, Storm HH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 7.0 (17.12.2014). Association of the Nordic Cancer Registries. Danish Cancer Society. Available from http://www.ancr.nu, accessed on 09/01/15.

### 1.1.5 Breast cancer risk factors

A number of breast cancer risk factors have been defined over the last century, including reproductive patterns and lifestyle factors [30, 31]. In this section, we will summarize changes in the frequency of some of the major risk factors for breast cancer during the last decades.

*Menopausal hormone therapy*

Steroid hormones play a key role in the development and progression of breast cancer, and menopausal hormone therapy is an established breast cancer risk factor [31]. The most reliable evidence for an increased risk of breast cancer caused by hormone therapy come from two randomized controlled trials[6] that compared breast cancer risk in women assigned to systemic estrogen treatment alone or in combination with progestin, to the risk in women receiving placebo. These studies indicated a 26-27% increased risk of breast cancer for women using combined low-dose estrogen and progestin preparation [32, 33], with the risk declining to levels in those untreated 2-3 years after treatment cessation [33].

The use of hormone therapy increased rapidly in Norway during the same period as the implementation of the Norwegian Breast Cancer Screening Program. During the late 1980s sale numbers were approximately 20 defined daily doses (DDD) per 1000 women, increasing to more than 90 DDD per 1000 women in the late 1990s for all

[6] *Randomized controlled trials (RCTs)* are experimental studies where two or more groups receive different interventions (for example treatments or health services) after random allocation of the study participants to each study group. RCTs are the highest level of evidence for effect in medical research.

products combined [34]. Following the publication of the results from the randomized trials in 2002 [32, 35], sale numbers dropped rapidly to 40 DDD per 1000 women in 2009 [34].

The Norwegian Women and Cancer (NOWAC) cohort contains information on hormone therapy use collected through questionnaires for the period 1996-2005 [36]. Random samples of Norwegian women were invited, and approximately 60% responded. These data indicate that in 1996, approximately 30% of NOWAC participants aged 48-62 years were current users of hormone therapy, increasing to 38% in 2002, and declining to 15% in 2005 [36, 37]. Nation-wide individual level data on hormone therapy prescriptions are available from 2004 and onwards in the Norwegian Prescription Database. According to these data, hormone therapy prescriptions are most frequent for women 55-59 years, and in this age group, systemic hormone preparations were prescribed to 23% of Norwegian women in 2004, declining to 12% in 2009 [34]. Among women who attended the Norwegian Breast Cancer Screening Program between 1996 and 2004, 43% had ever used hormone therapy [38].

Studies in Norway have generally resulted in stronger associations between use of hormone therapy and breast cancer risk than what was found in the Women's Health Initiative. In NOWAC, women who used estrogen and progestin in combination were found to have 150% higher breast cancer risk than never users, and a population attributable proportion[7] of 27% for women aged 45-64 years during the period 1996-2002 [37]. A similar attributable proportion was estimated for women aged 50-69 in 2002 in a nation-wide study using aggregated data [39]. Among women who attended the Norwegian Breast Cancer Screening Program between 1996 and 2004, ever use of hormone therapy was associated with 58% higher risk of breast cancer than never use, equivalent to a population attributable proportion of 20% among screening attendants [38].

Other studies from Scandinavian countries also support a stronger association than seen in the trials, and it has been suggested that this may be explained by higher doses of hormones in the preparations frequently used in Scandinavia [40]. High-dose preparations with estrogen and progestin has been the most common prescription group of systemic hormone therapy until the last part of the 2000s, when low-dose combinations became the main treatment [34]. Use of estrogen alone has not been demonstrated to increase the risk of breast cancer [41], but due to an increased risk of endometrial cancer such preparations have been recommended only for women who have had the uterus removed [42].

*Previous and current recommendations for follow-up of women who use hormone therapy*

Awareness of a possible increase in breast cancer risk associated with hormone therapy existed also prior the publications of the trial results in 2002. From the late 1970s until the mid-1990s, annual clinical breast examination was recommended for women using hormone therapy [42, 43]. In 1997, the first edition of the main Norwegian general practice textbook recommended mammography every second year for women using hormone therapy, and annual mammography for long-term users (> 10 years) and user with familial risk of breast cancer [44]. Currently, women who use hormone therapy are recommended mammography screening in the Norwegian Breast Cancer Screening Program according to clinical guidelines [45].

[7] *The population attributable proportion* is the proportion of disease in the population that can be attributed to the studied factor, and depends on both the excess risk among individuals exposed to the factor and the distribution of the factor in the population. The sum of attributable proportions for several different factors can exceed 100% since most, if not all, cases of disease have more than one cause.

*Other breast cancer risk factors*
Breast cancer risk increases with increasing age at first birth [46], weight gain in adulthood [47] and alcohol consumption [48], and decreases with increasing number of births [46]. It has been estimated that approximately 15% of the increase in breast cancer until the 1980s in Norway could be attributed to changes in childbearing patterns [49]. Since the 1970s, age at birth and the prevalence of obesity has increased steadily among Norwegian women [50, 51]. Alcohol consumption has increased over many decades [52]. Combined, changes in all these factors may have contributed to increasing incidence rates of breast cancer, as well as changes in other breast cancer risk factors.

## 1.2 Principles of mammography screening for breast cancer

### 1.2.1 The purpose of screening

Screening[8] refers to the process of identifying unrecognized disease in apparently healthy individuals [53]. The purpose of mammography screening is to detect breast cancer in a sufficiently early stage to improve its prognosis: compared to a situation without screening, where breast cancer would be detected as a consequence of symptoms, detection of pre-symptomatic cancers at screening will enable treatment in a more frequently curable stage [54]. In 1968, the World Health Organization described 10 principles for screening that would indicate the usefulness of a screening program for a specific disease [53]:

1) The condition sought should be an important health problem
2) There should be an accepted treatment for patients with recognized disease
3) Facilities for diagnosis and treatment should be available
4) There should be a recognizable latent or early symptomatic stage
5) There should be a suitable test or examination
6) The test should be acceptable to the population
7) The natural history of the condition, including development from latent to declared disease, should be adequately understood
8) There should be an agreed policy on whom to treat as patients
9) The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
10) Case-finding should be a continuing process and not a "once and for all" project

### 1.2.2 General definitions and concepts

The time between when a cancer is found through screening and the time it would have been found due to symptoms is termed lead time[9]. The time from when a cancer is detectable through screening and the time it would be detected due to symptoms is termed

[8] *Screening* is any examination that aims to detect unrecognized disease in apparently healthy individuals.
[9] *Lead time* is the time between the detection of breast cancer at screening and the time that the tumor would be detected if screening had not occurred, i.e. the amount of time that the date of diagnosis is advanced by screening.

sojourn time[10], which is the maximal lead time for a given tumor. Sojourn time is expected to vary for different tumors, in line with the heterogeneity of breast cancer [54]. The sojourn time is also dependent on how early in the natural history the screening test is able to detect the tumor. In practice, neither sojourn times nor lead times are directly observable, and their distributions are therefore largely unknown. Long sojourn times imply higher chances of being detected at screening [54].

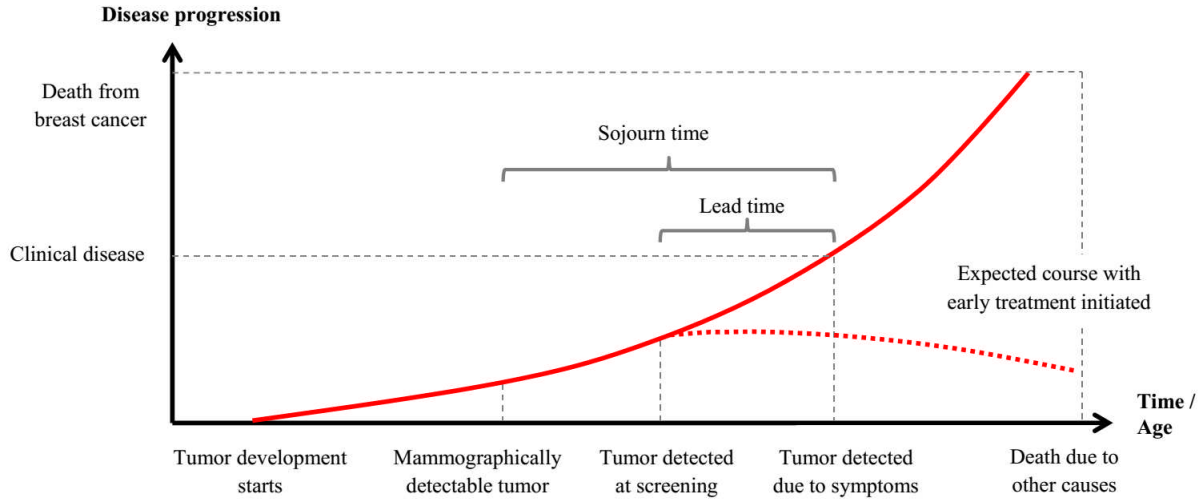



Figure 3. Principles of mammography screening

In mammography screening for breast cancer, the mammography examination in itself is not the final step in the diagnostic process. Women that are considered to have an abnormal mammogram are recalled for further investigation, and the diagnosis of cancer requires invasive procedures such as a biopsy. In consequence, the terms defined below do not refer to a single diagnostic test, but rather a sequence of diagnostic tests starting with mammography and proceeding to biopsy if necessary. Women who have an abnormal mammogram and who turn out to have breast cancer are classified as having *true positive* tests (‘*a*’ in the Table 1 below). An abnormal mammogram followed by normal subsequent tests is termed *false positive* (‘*b*’). When breast cancer is present, but is not detected at mammography, the mammography is classified as a *false negative* mammogram (‘*c*’). If there is no breast cancer and the mammogram is normal, the mammography examination is termed *true negative* (‘*d*’).

[10] *Sojourn time* is time from a breast cancer is detectable by the screening test to the time when the cancer would be detected in the absence of screening. Sojourn time is the maximum lead time.

Table 1. Measures of accuracy in diagnostic testing

| | | **Result of final diagnostic test** | | |
|---|---|---|---|---|
| | | Positive | Negative | |
| **Result of screening** | Positive | *a* | *b* | $PV+ = a / (a + b)$ |
| | Negative | *c* | *d* | $PV- = d / (c + d)$ |

$Sensitivity = a / (a + c)$
$Specificity = d / (b + d)$

The sensitivity[11] of mammography screening refers to the ability of mammography examination to correctly identify women who through subsequent tests turn out to have breast cancer, i.e. the proportion of true positives among the combined number of true positives and false negatives. The specificity[12] of mammography screening refers to the ability of mammography examination to correctly identify women without breast cancer, i.e. the proportion of true negatives among the combined number of true negatives and false positives. The predictive value of a positive test[13] (PV+) is the probability that a recalled woman will get a breast cancer diagnosis after subsequent tests, i.e. the number of true positives among all (true and false) positives.

The time between two screening examinations is termed the screening interval[14]. The European guidelines for screening programs defines interval cancers[15] as cancers that are detected in this time period in women who attended screening and had normal mammograms or normal recall investigations, or in a time period equal to the screening interval for women who have reached the upper age limit for screening [55]. As a group, interval cancers will consist of cancers that were missed at screening (false negatives) and cancers that were not detectable at screening, but reached a symptomatic stage before the next scheduled screening. The latter group will have short sojourn times and more aggressive growth patterns than cancers detected at screening [54].

The first screening round is often termed the prevalence screening[16]. The term may be used to describe the first screening in a population during screening implementation, or the first screening for an individual. When screening is implemented, women in the entire screening age range will have a prevalence screening. In a fully

[11] *The sensitivity of a test* is the ability of the test to correctly identify those individuals who have the disease that is being tested for. Sensitivity is calculated as the number of true positive tests divided by the number of true positive and false negative tests and expressed as a proportion or a percentage.
[12] *The specificity of a test* is the ability of the test to correctly identify those individuals who *do not* have the disease that is being tested for. Specificity is calculated as the number of true negative tests divided by the number of true negative and false positive tests and expressed as a proportion or a percentage.
[13] *The predictive value of a positive test* is the probability that an individual with a positive test truly has the disease that is being tested for.
[14] *The screening interval* is the time between two screening examinations. In the Norwegian Breast Cancer Screening Program, the screening interval is two years.
[15] *Interval cancers* are cancers that are detected during the screening interval in women who attended screening and had normal mammograms or normal recall investigations, or in a time period equal to the screening interval for women who have reached the upper age limit for screening.
[16] *Prevalence screening* is the first screening examination, either at a population or an individual level.

implemented screening program, the prevalence screening will primarily occur in the youngest age groups. The next screening rounds are termed subsequent or incident screening rounds[17].

### 1.2.3 Benefits and harms of mammography screening

The balance between benefits and harms of mammography screening has been debated for decades. Although the potential for side-effects of screening in general were discussed already in 1968 [53], the attention on such aspects in the specific situation of screening for breast cancer with mammography has increased over time [14, 54, 56]. The major harms of screening include overdiagnosis, false positive and false negative mammograms.

Overdiagnosis[18] of breast cancer in the context of a mammography screening program is defined as a breast cancer that would not be detected during the woman's lifetime in the absence of the program. One of the major challenges with overdiagnosis is that is it not possible to identify overdiagnosed tumors at an individual level. When a cancer is diagnosed at screening, it is not possible to foresee whether that particular cancer would progress to cause symptoms during the woman's remaining lifetime or not, since both the exact individual tumor progression and the woman's remaining lifetime is unknown at the time of diagnosis. As a result, all breast cancers are treated as potentially lethal. For women with tumors that would never be detected without the screening program, this treatment would be unnecessary, and would increase both the human and monetary costs associated with screening. In addition, a cancer diagnosis in itself may have a substantial impact on quality of life, even without considering adverse effects of treatment. A certain amount of overdiagnosis is inevitable in a screening program that succeeds in advancing the time of diagnosis to a preclinical stage.

Women who are recalled for further examination due to abnormal mammograms, but who do not have breast cancer, will be subjected to the various diagnostic procedures without experiencing any personal benefit, in addition to the mental distress associated with the cancer suspicion inherent in an abnormal mammogram. Women with interval cancer may be considered as given a false sense of security by the screening program.

In the context of a publicly financed mammography screening program, the costs of the program must also be considered in light of the balance between benefits and harms. For an individual woman, awareness of both the chance of preventing death from breast cancer and the risk of unnecessary diagnostic procedures and treatment, are important aspects in the decision to attend or not to attend when invited for mammography screening.

[17] Incident or subsequent screening refers to all screening rounds or screening examinations after the prevalence screening.

[18] Overdiagnosis due to mammography screening is the detection of breast cancer at screening that would not have caused symptoms during the woman's lifetime, and thus would not have been detected without screening.

## 1.3 Screening for breast cancer in Norway

### 1.3.1 The Norwegian Breast Cancer Screening Program (NBCSP)

*Background for implementation of the NBCSP*[19]

In November 1985, the Director General of Health[20] appointed a working group to make a professional evaluation of the literature and experiences on mass screening of breast cancer [57, 1.2, page 7]. The Norwegian Health Authorities' request was prompted by publications of results from randomized controlled studies indicating an improved total survival [57, 1.1, page 7], and especially the results from a Swedish trial in 1985 on screening for breast cancer. Also the decision from the Swedish National Board of Health and Welfare to introduce breast cancer screening, urged a Norwegian investigation [57, 1.1, page 7]. The mandate was extensive and included besides the feasibility and benefits of breast cancer screening, an evaluation of the administrative and educational challenges, and of the costs and radiation risks. The working group consisted of members from the Norwegian Radium Hospital, the Ullevål Hospital, the Norwegian Directorate of Health[21], the National Institute of Radiation Hygiene[22], and the Cancer Registry of Norway. The official Norwegian report was handed over in December 1986 [57].

The group recognized that the high incidence of breast cancer in Norway, with a high mortality, and that the treatment regimens used the last 10-20 years had not changed the survival compared to the 1950s, were important arguments for considering screening to detect cancer in an earlier stage of the disease.

Based on a thorough review of the existing literature, and particularly on the Swedish WE trial [58], the group concluded that there was an undisputable reduction in breast cancer mortality among women randomized to screening, aged 50-74 years old [57, 5.5, page 23]. After discussing benefits and harms, such as interval cancer and overdiagnosis, costs, and mandatory requirements for a national screening program, they recommended to start a national program for women aged 50-74 years. Under some doubts they also recommended a similar program for women aged 40-49.

The Norwegian government decided to start a national program for women aged 50-69 in 1994 and allocated NOK 22.7 mill [59]).

*Implementation and organization of the NBCSP*

Population-based mammography screening in Norway started as a pilot project in four Norwegian counties in 1995/96, as a collaboration between the Cancer Registry of Norway, the National Health Screening Service[23], the Norwegian Radiation Protection Agency[24], and the regional authorities in the four counties Rogaland, Hordaland, Oslo and Akershus. The target age group was women 50-69 years (born 1927-46), and a pilot period of 4 years with 2 year screening intervals followed by evaluation was planned. However, in 1998 the Storting[25] approved nationwide extension of the program, and the Norwegian Breast Cancer Screening Program was established. The counties Telemark,

[19] The terms used in this section refer to the terms as they were used in the cited reference NOU 1987:7 [57] and may not correspond to the terms used today.
[20] Norwegian name: Helsedirektøren
[21] Norwegian name: Helsedirektoratet
[22] Norwegian name: Statens institutt for strålehygiene
[23] Norwegian name: Statens helseundersøkelser
[24] Norwegian name: Statens strålevern, formerly Statens institutt for strålehygiene
[25] The Norwegian parliament

Aust- and Vest-Agder were included in the NBCSP in 1999, Troms and Finnmark in 2000, Østfold, Nordland, Buskerud and Nord- and Sør-Trøndelag in 2001, Oppland and Møre og Romsdal in 2002, Sogn og Fjordane in 2003, and Vestfold and Hedmark in 2004. By the end of 2005, the program was fully implemented and all women in the target age group had received at least one invitation for mammography screening.

Women are invited according to birth cohort, and all women who will be within the target age group during each 2-year screening round are invited. This implies that women may be between 48 and 73 years at invitation, depending on the woman's birth month and the time of year a new screening round starts in each county [29]. Each woman receives a maximum of 10 invitations. Invitations are organized according to municipality of residence.

A screening mammography examination consists of two-view mammograms (craniocaudal and mediolateral oblique view) and interpretation by two radiologists (independent double reading). Mammograms that are interpreted as possibly abnormal[26] by one or both radiologists are discussed at consensus or arbitration meetings, and a decision on whether or not to recall the woman for further examinations is made. During the first period of screening implementation, screen film mammography was used, but from 2000 there has been a gradual transition to full field digital mammography [60].

Before the implementation of the NBCSP in each county, multidisciplinary breast diagnostic centers responsible for interpretation of the mammography examinations, diagnostic work-up and breast cancer treatment were established. Currently there are 16 breast diagnostic centers with a total of 27 stationary and mobile (buses) screening units. The multidisciplinary teams associated with each center consist of radiologists, radiographers, nurses, pathologists, surgeons and oncologists. Weekly meetings with discussion and planning of management of breast cancer patients are recommended. The establishment of breast diagnostic centers with multidisciplinary teams has contributed to the centralization of breast cancer diagnostics and treatment in Norway.

The Ministry of Health and Care Services has the overall responsibility for the NBCSP. The Norwegian Directorate of Health is responsible for the health care quality of the program, whereas the Cancer Registry of Norway has the administrative and operational responsibility. The Norwegian Institute of Public Health has administered the invitation letters and result letters. The Norwegian Radiation Protection Agency has been responsible for the technical quality control of the radiologic equipment and controlling the radiation doses. The regional health authorities are responsible for the breast diagnostic centers conducting the examinations and for follow-up of the results (diagnostic work-up and treatment). Information on attendance, screening outcomes and recall examinations are reported electronically to Cancer Registry of Norway through a closed data network. The NBCSP has its own quality assurance manual, based on the European guidelines for quality assurance in breast cancer screening and diagnosis [55]. Representatives of professional groups involved in mammography screening form an advisory board that monitors and offers advice on quality assurance functions and modifications of guidelines or program processes.

In each screening round, approximately 76% of the invited women attend, whereas overall, 83% of the invited women attend one or more screening examinations [61]. During the first 10 years of the program, 4.6% of women attending their first screening were recalled for further diagnostic tests. At subsequent screening, 2.6% were

[26] The mammograms are scored according to a scoring system ranging from 1 to 5, and all mammograms with a score of 2 or more from either or both radiologists are discussed at consensus meetings.

recalled [62]. The detection rate for invasive breast cancer and DCIS combined was 6.5 per 1000 prevalent screens, and 4.9 per 1000 subsequent screens. The positive predictive value of a recall examination was 13.8% and 18.7% in prevalent and subsequent screenings, respectively. When considering all interval cancers as false negative, sensitivity was estimated to 76.4% and specificity to 96.4% for prevalent and subsequent screenings combined [62]. From 1996 to 2007, breast cancers detected at screening comprised 67% of all breast cancers detected among invited women aged 50-69 years [61].

In 2012-2013, 550 000 women were invited, and 410 000 (74%) attended. 11 400 were recalled for further examinations (2.8%), and 2160 women were diagnosed with invasive breast cancer or DCIS during recall examinations, which gives a predictive value of a positive mammogram of 18.9% [63].

### 1.3.2 Breast cancer screening for women with familial breast cancer risk

In Norway, the small group of women with a high familial risk of breast cancer have been offered more intensive follow-up than the public screening program [11, 26]. Currently, women with familial risk without known germline mutations in high-penetrance genes[27] are recommended annual mammography every year from age 30 to age 60, and NBCSP participation from age 60 years [11]. Women with known germline mutations in high-penetrance genes are recommended annual MRI examination from 25 years of age. *BRCA1/2* mutations are the most frequent of these [64], with substantial variation in the prevalence of each mutation across municipalities in Norway [65]. Norwegian women who carry *BRCA1* mutations have a 60% cumulative risk of developing breast cancer [66] and a similar risk of developing ovarian cancer [67]. Most women with identified mutations choose to have their ovaries removed before age 40 years, whereas preventive mastectomy has been less frequent [64], but increasingly used during the last decade [68]. Annual examinations including mammography has been recommended according to Norwegian guidelines for women with familial risk of breast cancer since 1994 [26], and genetic tests were developed during the last part of the 1990s [69].

Follow-up for women with familial risk of breast cancer is organized by the departments of medical genetics[28] within the regional health authorities [64], and not by the NBCSP. According to the NBCSP quality assurance manual, women who have separate follow-up for familial breast cancer should be registered as opted out of the NBCSP [70].

### 1.3.3 Non-organized breast cancer screening in Norway

There are few available sources on the extent of mammography use outside the NBCSP in Norway, most of which have not been able to reliably distinguish clinical mammography (i.e. prompted by symptoms) from organized screening of women at high risk (see section

[27] *High-penetrance genes* include *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *CDH1* and *STK11*

[28] Current *criteria for referral* to a department of medical genetics for genetic counselling and assessment of familial risk of breast cancer including testing for *BRCA1/2* are: breast cancer before age 50 or ovarian cancer before age 70 years, breast cancer and ovarian cancer, breast cancer in both breasts at the time of diagnosis, detected mutation in *BRCA1/2* in a close relative, one or more first degree relatives with breast cancer before age 50 years, several breast cancer cases in the family, both breast and ovarian cancer in the family, and several cases of breast or ovarian cancer in second or third degree paternal relatives.

1.3.2) or from non-organized screening (also termed opportunistic screening)[29]. By non-organized screening, we refer to mammography examination of women who are free from breast cancer symptoms, at the woman's own initiative, as opposed to receiving an invitation from a defined screening program.

The first report on mammography use in Norway from the Norwegian Radiation Protection Agency shows that in 1983, mammography examinations were rare or not available in most counties in Norway [71]. The total number of examinations was 10 481 (2.5 per 1000 inhabitants), and more than half of these were conducted in Oslo, and some in Troms, Sør-Trøndelag and Sogn og Fjordane. In 1988, the total number of mammography examinations had increased to 77 128 (18.3 per 1000 inhabitants) [72] and in 1993, the number of mammography examinations was approximately 220 000 (41.6 per 1000 inhabitants), with 70% of the examinations conducted at private institutions and higher rates in urban than rural areas. A quality control in 1993-1994 concluded that there were large variations in the quality of both the mammography equipment and the examination procedures [73]. All the numbers above refer to the number of examined breasts, and not the number of examined women. In 2002, the total number of examinations was 349 056 (76.7 per 1000 inhabitants) and 36% of these examinations were conducted at private institutions. The number from 2002 consists of a combination of reporting both breasts as one examination and reporting of each breast as a separate examination, due to insufficiently detailed information from the radiology institutions [74].

According to a study of mammography use at a private radiology clinic in Kristiansand in 1992, most of the examinations were conducted among asymptomatic women referred from a doctor [75]. The majority of women examined were aged 45-60 years. Fifty-two percent had previously been examined, but the number of women who had regular examinations corresponded to no more than 1.2% of women aged 40-70 in Aust-Agder and Vest-Agder combined. Regular mammography was most common among women 40-55 years. A greater proportion than expected lived in municipalities near the clinic, indicating that availability increased the examination frequency. In 1991, the clinic conducted 69% of the mammography examinations in the two counties. In total, 16% of all women in the two counties were examined in 1991. The corresponding national numbers were 11% (75 000 women) [75].

Since the start of NBCSP implementation in 1995/96, a questionnaire on breast cancer risk factors and previous experience with mammography has been sent along with the first screening invitation. Data from these questionnaires indicate that 64% of screening attendants had at least one mammography examination prior to their first NBCSP attendance, and 38.5% had been examined with mammography no more than 3 years before their first program screening [76].

In 2003-2004, the Cancer Registry of Norway conducted a pilot study to assess the feasibility of including information in the screening database on mammography examinations performed outside NBCSP [77]. Twenty-one private clinics offered mammography in 2003 and examined in total 86 370 women of all ages. Most performed mammography both with and without referral from a doctor. The majority of clinics recommended their "screening costumers" annual examinations. In addition, detailed data from two private radiology clinics in Hedmark and Vestfold were collected over a period of 4 months before these counties were included in the NBCSP. The clinic in Hedmark

[29] *Non-organized or opportunistic screening* is examination of apparently healthy individuals at the individual's own or his/her doctor's initiative.

performed screening mammography only and examined 2123 women in total and 1290 (66%) in the 50-69 years age group, corresponding to 17% of NBCSP target population in the county. Five cases of breast cancer were detected during the data collection period. The clinic in Vestfold performed both screening and diagnostic mammography and examined 1041 women in total and 556 (56%) in the 50-69 years age group, corresponding to 7% of NBCSP target population in the county. Nine cases of breast cancer were detected during the data collection period.

Following an assignment from the Norwegian Cancer Society, Hofvind and Sanderud collected information on the use of mammography in Norway [78]. The authors collected data for 2005 and 2008 from hospitals, private clinics, the Norwegian Directorate of Health and the Cancer Registry of Norway. Private clinics offering mammography were available in all counties except Finnmark, Nord-Trøndelag and Sogn og Fjordane in the two study years. The counties with the highest proportion of mammography in private clinics were Oslo, Møre og Romsdal, Hedmark, Nordland and Akershus. On a nationwide basis, 8-10% of women 40-49 years and 6% of women 50-69 and 70-75 years had mammography in private clinics in 2005 and 2008. In approximately half of these examinations the women were referred from a doctor.

In 2010, 7000 women aged 45-55 years were invited to participate in a questionnaire-based survey conducted by SINTEF on use of mammography [79]. Fifty percent of the invited women replied. Among women aged 45-49, 54% had been examined with mammography, compared to 88% of women aged 50-55 years. In the youngest age group, the majority was referred from a doctor, and the most common causes for examination were symptoms, worries about breast cancer, familial risk or recently detected breast cancer among family or friends. In the oldest age group, invitation from the NBCSP was the main reason for examination.

## 1.4 Previous Norwegian evaluations of mammography screening

### 1.4.1 Background

During the period 1963-1982 several studies were organized internationally to study screening mammography as a mean to reduce breast cancer mortality on a population level. The studies were designed as randomized controlled trials (RCTs) with a control group not offered screening. They varied with respect to scientific design and the age groups studied. Breast cancer mortality reduction reported from these and other studies varied, which was one of the reasons that a still on-going discussion about the applicability of mammography screening as a public health intervention started.

### 1.4.2 Norwegian technology assessment

Also in Norway, a debate started on the background of the international assessments, first by a letter from the National Committee for Medical and Health Research Ethics[30] to the Minister of Health Affairs in 2001.

This was followed up among clinical experts, as well as in the Ministry of Health and Social Affairs[31] and its National Council on Priorities in Health Care[32], and led to a

[30] Norwegian name: Den nasjonale forskningsetiske komité for medisin og helsefag (NEM)

request from the Norwegian Directorate for Health and Social Services[33] to the Norwegian Centre for Health Technology Assessment[34] to assess the evidence of the benefits (and harms) of routine mammography screening.

The assessment was based on already published systematic reviews of the seven international RCTs (secondary literature), since there was general agreement internationally with respect to which of the RCTs to include in the reviews. The disagreement in the published systematic reviews of these RCTs referred to the assessment of their scientific quality, and the interpretation of their results.

The report identified 15 systematic reviews published 1995-2002, all of which were assessed and evaluated according to variation in screening age range, follow-up, and criteria for assessing the RCTs. The report also included some results from national population based screening programs published from 1995 and later.

The report was developed by an internal research group at the Norwegian Centre for Health Technology Assessment and evaluated by Nordic experts in cancer treatment and epidemiology [80].

The following positive health effects were summarized from 15 identified systematic reviews of seven RCTs:

*Reduction of mortality, age groups 50-69 and 40-49 years*

For the age group of 50-69 years, the assessment showed a reduction in risk of breast cancer mortality. The scientific documentation had some weaknesses and the estimated reduction in breast cancer mortality was in the order of 6-27%. All the systematic reviews indicated clear reductions, with the exception of the Cochrane report (see below). The Cochrane report gave the lowest risk reduction (6% in trials of medium quality). For the age group 40-49 years the results showed no significant effect of screening on mortality. Few of the included systematic reviews had any information of total mortality. In a similar report from the Dutch Health Council, "The benefit of population screening for breast cancer with mammography", the risk ratio (RR)[35] for death from all causes among women 50 years and older was estimated to 0.99 (95% CI 0.97 to 1.02) after 13 years .

The following negative health effects were summarized from results of population based screening programs:

*False positive mammograms*

Based on figures from screening programs in England and Denmark, 10-14% of women will have a false positive mammography test after three rounds of screening. This proportion was expected to increase with each screening round

[31] Norwegian name: Helse- og sosialdepartementet
[32] Norwegian name: Prioriteringsutvalget
[33] Norwegian name: Sosial- og helsedirektoratet
[34] Norwegian name: Senter for medisinsk metodevurdering (SMM)
[35] *Risk ratio* is a measure of comparison of disease occurrence and is calculated as the risk of a particular event in one study group divided by the risk in another group

*DCIS and overdiagnosis*
DCIS constituted 11-20% of all cases of breast cancers diagnosed at the first and following screenings. The report did not identify systematic reviews that specifically referred to overdiagnosis.

*Interval cancers*
At least one fifth of all cases of breast cancer in the screened group were diagnosed between screenings as interval cancers.

The report emphasized that the women in the relevant age groups must be informed about all aspects of mammography screening as a health offer, the benefit as well as the negative effects of screening.

### 1.4.3 Report "Mammografiscreening av kvinner 40-49 år" from the Norwegian Knowledge Centre for the Health Services, 2007.

The Norwegian Knowledge Centre for the Health Services[36] was established in 2004, and included among other scientific institutions also the Norwegian Centre for Health Technology Assessment. On its own initiative, the Knowledge Centre summarized the present knowledge on benefits and harms of offering routine mammography screening to women age 40-49 years. The included studies were three systematic reviews from 2002, and one RCT from the United Kingdom published in 2006. The report summarized the reduction in breast cancer mortality to 16% (95% CI 4 to 27%), corresponding to an absolute risk reduction of 0.0003, or one prevented breast cancer death per 3000 invited to screening.

## 1.5 Selected previous international evaluations of mammography screening

### 1.5.1 Reviews and meta-analyses of randomized trials

**Screening for breast cancer with mammography**
Gøtzsche PC, Jørgensen KJ.
Cochrane Database Syst Rev. 2013 Jun 4;6:CD001877.

The first review on the effect of mammography screening on breast cancer mortality from the Cochrane Collaboration was published in 2001, and has been updated several times, most recently in 2013. The most recent report is based on results from seven randomized trials comparing mammography screening with no mammography screening. The included studies comprised more than 600 000 women aged 39-74 years. In the latest update, the effects of mammography screening on breast cancer incidence and treatment are also addressed. Among the seven included trials, three were considered as adequately randomized with a low risk of bias. The remaining studies were considered to carry a high risk of bias due to cluster randomization and/or inconsistencies in the description of

[36] Norwegian name: Nasjonalt kunnskapssenter for helsetjenesten

randomization procedures, inclusions and exclusions across publications, and also lack of blinding in assessment of cause of death. In most of the trials, women in the control group were offered screening at the end of the trial period.

When results for women from all included age groups were combined, the risk ratio for death from breast cancer was 0.81 (95% CI 0.74 to 0.87) in favor of screening for all seven trials and 0.90 (95% CI 0.79 to 1.02) when only the trial deemed as adequately randomized were considered.

For women 50 years or older, the risk ratio for death from breast cancer specific after 7 years of follow-up was 0.72 (95% CI 0.62 to 0.85) in favor of screening. When restricted to the trials considered as adequately randomized, the effect was 0.88 (95% CI 0.64 to 1.20). After 13 years of follow-up, the risk ratio for death from breast cancer was 0.77 (95% CI 0.69 to 0.86) for all trials combined and 0.94 (95% CI 0.77 to 1.15) for the trials considered as adequately randomized.

For deaths from all causes, there were no clear differences between the screened groups and the control groups.

The screened groups had more breast surgery (RR 1.31, 95% CI 1.22 to 1.42) and more radiation therapy than the control groups (reported only in two trials). In the trials considered as adequately randomized, the number of breast cancer diagnoses was 25% higher in the groups offered screening (95% CI 18 to 34%) than in the control groups after 7-9 years follow-up. In the trials considered as sub-optimally randomized, the number of breast cancer diagnoses in the screening groups was 33% higher (95% CI 24 to 44%) than in the control groups before the control groups were offered screening.

The authors concluded that if screening reduces mortality from breast cancer by 15% and results in 30% overdiagnosis, 2000 women would need to be invited for screening during 10 years to prevent one breast cancer death and 10 women would be diagnosed with and treated for breast cancer that would not have been detected without screening.

**The benefits and harms of breast cancer screening: an independent review.**
Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M.
Br J Cancer. 2013 Jun 11;108(11):2205-40.

The Independent UK Panel on Breast Cancer Screening was convened by the Cancer Research UK and the Department of Health in England to evaluate the benefits and harms in the context of the UK mammography screening programs. The programs invite women aged 50-70 years every three years.

The Panel's evaluation was based on the same randomized trials as those included in the Cochrane review. The combined estimate from these trials after 13 years of follow-up indicated a 20% reduction (95% CI 11 to 27%) in breast cancer mortality in women invited for screening. The Panel emphasized the uncertainty in this estimate due to sources of bias in the trials and the differences between the trials, conducted in the 1970s and 1980s, and the current screening programs. The absolute mortality benefit was estimated to be one breast cancer death prevented for every 235 women invited to screening. This estimate was obtained by applying the 20% risk reduction from the trials to the cumulative absolute risk of death from breast cancer for women aged 55-79 years in the UK.

Overdiagnosis was considered as a major harm of mammography screening and was estimated from the randomized trial that did not offer screening to women in the

control groups at the end of the trial. The number of excess cancers (overdiagnosis) was estimated as the difference in cumulative numbers of incident breast cancers in women invited or not invited to screening, using the longest available follow-up period. The Panel estimated the following measures of overdiagnosis:

A. Excess cancers as a proportion of cancers diagnosed over the whole follow-up period in unscreened women
B. Excess cancers as a proportion of cancers diagnosed over the whole follow-up period in women invited for screening
C. Excess cancers as a proportion of cancers diagnosed during the screening period in women invited for screening
D. Excess cancers as a proportion of cancers detected at screening in women invited for screening

The Panel's estimate of overdiagnosis from a population perspective (method B), i.e. as a proportion of cancers diagnosed from the start of the screening period and the throughout the rest of the women's lives, was 11%. From an individual perspective (method C), i.e. the probability that a cancer diagnosed during the screening period is overdiagnosed, was estimated to 19%. Both estimates include both invasive cancer and DCIS. The Panel emphasized that the uncertainty in these estimates are even greater than for the estimates of reduction in mortality, due to the lack of data available to answer this question directly.

The Panel concludes that for every 10 000 UK women invited to screening, 43 deaths from breast cancer will be prevented and 129 cancer diagnoses will represent overdiagnosis.

### 1.5.2 Evaluations of population-based mammography screening programs

The EUROSCREEN Working Group presented the results from their evaluation are in an entire Supplement Issue of Journal of Medical Screening (J Med Screen 2012; 19 Suppl 1). The summary below is based primarily on the summary report in that issue:

**Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet.**
Paci E; EUROSCREEN Working Group.
J Med Screen. 2012;19 Suppl 1:5-13.

The EUROSCREEN Working Group evaluated current European population-based mammographic screening programs using observational data. Most of the included studies were from countries with programs offering biennial screening to women aged 50-69 years.

Mortality reduction was estimated separately for incidence-based mortality[37] (cohort) studies[38], case-control studies[39] and trend studies[40]. The combined estimate of

[37] *Incidence-based mortality* or *refined mortality* from breast cancer is a mortality rate that counts only the breast cancer deaths occurring among women who had their cancer detected after a specific time point, such as after screening invitation. See also chapter 4.2 for a further description.
[38] *Cohort studies* follow groups of individuals who have or do not have specific characteristics (termed exposure) and investigate whether one group is more or less likely to develop a specific disease or condition (termed outcome) when followed over time.
[39] *Case-control studies* investigate whether a specific characteristic (exposure) is more or less frequent among groups of individual who have or do not have a specific diseases or condition (outcome).

breast cancer specific mortality from seven incidence-based mortality studies was RR 0.75 (95% CI 0.69 to 0.81) for women invited to screening compared with women not invited to screening. The eight case-control studies included also indicated a reduction in risk of death from breast cancer (OR 0.69, 95% CI 0.57 to 0.83) for women invited to screening. In trend studies estimating the annual percent change in breast cancer mortality, reductions ranged from 1% to 9% per year for studies with follow-up at least 10 years from program implementation. Other trend studies compared breast cancer mortality in time periods within a country and reported breast cancer mortality 28% to 36% lower in the screening periods compared with prescreening periods.

Overdiagnosis was estimated as a proportion of the expected incidence in the absence of screening, but with variation in the age range of the denominator and in whether DCIS was included in the estimate. Estimates ranged from 0 to 54% with no or suboptimal consideration of underlying breast cancer incidence and lead time. In the studies considered most reliable by the authors, estimates ranged from 1 to 10% after accounting for underlying incidence and lead time.

The cumulative proportion of women who are recalled for further examinations was estimated to 20% during 10 screening rounds. Among these, 3% had invasive tests.

The authors concluded that for every 1000 women screened (i.e. attending, not invited as in the Cochrane and the UK Panel reviews) 71 breast cancers would be detected, seven to nine deaths from breast cancer would be prevented, and four women would be overdiagnosed. In addition, 200 women would be recalled for further assessment.

**Dépistage systématique par mammographie. Rapport du 15 décembre 2013.**
Swiss Medical Board, Organe Scientific
Available from http://www.medical-board.ch/

The Swiss Medical Board is an independent organ formed by the Health directors of the Swiss Cantons, the Swiss Medical Association, the Swiss Academy of Medical Sciences, and the Government of Lichtenstein, with the purpose of evaluating health care intervention financed through the mandatory care insurance, in particular with respect to cost-effectiveness.

There is no national mammography screening program in Switzerland, but a proportion of the cantons had in 2013, when the report was written, already implemented a screening program or were planning to do so. The programs have a slightly varying target age group (50-69, 50-70 and 50-74 years) with a screening interval of 2 years.

The authors conducted a literature search for meta-analyses[41] of RCTs. They also included meta-analyses of observational studies. They concluded that the RCTs support a reduction in breast cancer mortality due to mammography screening, but no reduction in total mortality, and that the observational studies indicate a smaller effect of screening than the RCTs.

In the analyses of cost-effectiveness, the results for mortality and overdiagnosis from the Cochrane review from 2011, based on 13 years of follow-up for women aged 50 years or older, were applied, with 16 deaths from breast cancer avoided and 35

[40] *Trend studies* investigate changes in the occurrence of a specific disease or condition over time, using aggregated data.
[41] *A meta-analysis* is a summary analysis of multiple studies, including statistical analyses that combine the results across studies.

overdiagnoses per 10 000 women invited for screening. The probability of false-positive examinations was set to 4%, based on the EUROSCREEN publication by Hofvind et al [81], corresponding to 1025 per 10 000 women invited for screening for 6.5 years. Costs were calculated based on reimbursement fees for screening and breast cancer treatment and were hence the only Swiss data included in the cost-effectiveness estimation. Only direct costs were included. Quality-adjusted life-years[42] were calculated based on the Karnofsky index, which is the standard method used by the Swiss Medical Board [82]. The analyses were presented for 10 000 women invited for screening for 6.5 years (three screening rounds), and followed for a total of 13 years.

The number of quality-adjusted life-years was similar in the two groups (invited and not invited for screening). Costs of screening were estimated to CHF 810 per woman invited for three screening rounds, including breast cancer treatment of additional 35 patients per 10 000 women examined. The Swiss Medical Board concluded that no quality adjusted life years are gained by mammography screening, and that the mammography screening programs in the Swiss cantons should not be sustained.

[42] *Quality-adjusted life-years* is a measure of disease burden, including both the quality and the quantity of life lived

# 2 The evaluation assignment

In 2006, the Ministry of Health and Care Services charged the Research Council of Norway with responsibility for conducting a research-based evaluation of the Norwegian Breast Cancer Screening Program. Report no. 1 (2006-2007) to the Storting states that "there is a need to evaluate the extent to which the Norwegian Breast Cancer Screening Program has fulfilled its intentions and purpose, and to establish a scientific basis for potential expansion of the screening program to include other age groups."

Special focus was to be placed on whether the Norwegian Breast Cancer Screening Program has attained its primary target of achieving a 30% reduction in breast cancer-related mortality among women invited to take part in the screening.

## 2.1 Cooperation agreement

The cooperation agreement between the Ministry of Health and Care Services and the Research Council of Norway describes the relation between the Ministry and the Research Council in relation to the performance of the research-based evaluation.

Overall, the agreement fulfils the following purposes:

- Facilitate research-based evaluation of the high scientific quality
- Maintain conflicts of interest by ensuring the necessary distance between the Ministry of Health and Care Services as the ministry responsible for implementation of the program and the evaluation of the program
- Develop The Research Council's advisory role to the Ministry and strengthen collaboration on research-based evaluation
- Improve utilization of research results in the follow-up of a national health service deal
- Contribute to the development of health services research and enhance knowledge and skills in research-based evaluation
- Exploit the Research Council's competence and network

The Ministry of Health and Care Services was responsible for:

- Ensure that the document describing the objectives and framework, in dialogue with the Research Council of Norway, is in line with the assignment given by the Ministry

The role of the Research Council of Norway was to:

- Develop the objective and framework document
- Maintain evaluation mission in line with the objective and framework document
- Implement an appropriate structure to ensure the technical and administrative support to the assignment
- Ensure the scientific and methodological quality by selecting projects and be responsible for project implementation and follow-up
- Maintain contact with the research community

- Report and advise the Ministry about project status, progress and conclusions
- Establish appropriate procedures for the dissemination of the results of the evaluation

The cooperation agreement was originally valid until 31.12.10 but was, due to delays in the evaluation, prolonged till 31.12.14.

The Ministry of Health and Care Services allocated 18.2 million NOK in total to the evaluation.

## 2.2 Steering Committee

A steering committee was appointed by the Research Council of Norway. The task of the steering committee was the following:

- to ensure the scientific follow-up of the research-based evaluation of the Norwegian Breast Cancer Screening Program
- to prepare an objective and framework document for the evaluation in cooperation with the Research Council of Norway
- to allocate research funds to the evaluations projects
- to prepare and approve the final report to the Ministry of Health and Care Services

The members of the steering committee provided declarations on conflicts of interest which are available from the following website (in Norwegian):
http://www.forskningsradet.no/prognett-mammografi/Styringsgruppen/1226994052806

## 2.3 Objectives and framework

The objective and framework document (see Appendix I) clarifies the purpose of the evaluation and specifies the research topics to be addressed in the evaluation.

The evaluation aimed to address three main topics:

- Evaluation of effectiveness of the screening program on mortality due to breast cancer, changes in staging, and changes in the incidence of advanced cancer
- Evaluation of the organization, availability and quality of the screening program as well as associated scientific development
- Economic evaluation: analysis of the combined use of resources and the benefit/effectiveness of the screening program

For each of the main topics there were formulated indicators with current issues.

## 2.4 Selection and follow-up of evaluations projects

### 2.4.1 Call for proposals

The Research Council of Norway invited the research community to forward project applications within the deadline 16th of April 2008. Norwegian as well as foreign institutions was eligible as project owner. Due to considerations of impartiality, persons employed in the Cancer Registry of Norway could not be part in applications for projects.

The call endorsed the evaluation mission as specified in the objective and framework document, and stated that projects studying the effectiveness of the program in terms of changes in disease burden of breast cancer, changes in mortality due to breast cancer, changes in staging, and occurrence of interval cancers were particularly wanted. The variety of studies of the program's effectiveness could range from the sensitivity and specificity of mammography screening to patient experiences. Studies of the organization, availability, and quality of the screening program, as well as professional development related to the program, were also of interest. In addition, projects that involved financial evaluations of the program and the program's effectiveness were desired.

The call was published on the Research Council of Norway's website. In addition information was sent to NOS-M (the Joint Committee of the Nordic Medical Research Councils), NIH (National Institute of Health, US) Inserm (France), IARC (International Agency for Research on Cancer, France), MRC (Medical Research Council, UK) and ESF (European Science Foundation) with a request to forward the call in the way each institution found appropriate.

### 2.4.2 Application process

Eleven applications were received within the deadline. Eight of the applications had Norwegian project owners, the remaining three applications had, respectively, Swedish, English and Dutch project owners.

Total funds applied for were around 45 million NOK. The applications were processed by an international expert panel (Norwegian, English, Finnish and Dutch member). The expert panel evaluated the project's scientific quality only.

On the basis of the applications, the reviews from the expert panel and relevance to the evaluation mission, the steering committee granted funding to seven projects. In the application process, the steering committee emphasized the Ministry's desire of an evaluation with special focus on mortality. The steering committee also found it important that the prioritized research questions were investigated from different perspectives and with different methods to provide a broad evidence basis. The steering group thus decided to invite at least two different research groups to evaluate mortality reduction, over-diagnosis and health economy.

The funded projects were related to the following themes in the objective and framework document:

- Changes in breast cancer mortality
- Interval cancer occurrence
- Extent of overdiagnosis
- Women's experiences
- Economic evaluation and cost-effectiveness analysis

The funded projects were asked to submit revised applications based on the funding allocated to the specific project and the elements to be prioritized in the application.

Evaluation of the organization, availability, quality, and associated scientific development of the screening program was not covered by the funded projects. Some of these elements were however included in the proposals that did not meet the desired quality. Due to limited resources it was decided not to proceed with these applications.

### 2.4.3 Progress reporting

The seven evaluation projects have been followed through progress reporting, meetings and seminars with the researcher groups. All projects have delivered final reports and all publications resulting from the evaluation projects have been sent to the steering committee. The steering committee has requested additional information from the project groups when needed.

## 2.5 Creating the databases for the evaluation projects

The call for proposals mandated that the evaluations should be based on individual data, from a compact database and with an identical set of data for project groups with the same evaluation contract.

Meetings with the Cancer Registry disclosed that the existing database of the Norwegian Breast Cancer Screening Program, designed primarily for administrative purposes for the four pilot counties, did not meet the requirements for extended quality control, linkage to other registries and research. The Cancer Registry agreed therefore to develop a new project database for this specific purpose based on the relevant data stored in the various databases in the Cancer Registry (see Appendix II). The process was embodied in a Project Directive commissioned by the Director of the Cancer Registry and a steering group with two members from the Cancer Registry and one from the steering committee was appointed to monitor the work.

The main challenges included to create new, documented and quality controlled data files combining data from the screening database and the incidence database, to transform the screening database files from the level of screening rounds to the individual level, to link the two files on individual level, and to describe and document all the variables in the new database. The project base was established in May 2009 with a comprehensive description of all variables, including four different sets of data: identifiers for all women, information on screening procedures (date of invitation, attendance, results of the screening with dates for all events), information on all lesions (all invasive cancers from 1953 through 2009, all DCIS from 1993 through 2009, all hyperplasia/atypia diagnosed in the screening program. Prognostic information on tumor characteristics and date of diagnoses were also included), and questionnaires on risk factors for breast cancer from the first screening round.

In 2001 a set of new regulations applicable for the Cancer Registry entered into force, and in 2009, the Data Protection Authority pointed out that the Norwegian Breast Cancer Screening Program was not consistent with the applicable regulations as expressed informed consents could not be documented. At the same time, the Cancer Registry was required to store these same data on behalf of the health institutions, in order for these institutions to fulfil their legal obligation to keep medical records, and thus the Cancer Registry could not erase these data. The data could be used for clinical purposes,

but not for research or evaluation. This caused a considerable delay in finalizing the data for the evaluation.

After several rounds of discussions the Cancer Registry agreed in 2011 to assist the Research Council in the application process on behalf of the research project groups. In December 2011 the Cancer Registry obtained permission from the Regional Committee for Medical and Health Research Ethics and the Norwegian Data Protection Authority[43] to process and link the screening data stored on behalf of the health institutions with the national health registries as well as Statistics Norway for evaluation and research purposes related to the NBCSP and breast cancer. The data files including data from the Cancer Registry were available for the various project groups in January 2012.

Linkage with the other registries required independent application to each separate registry, and a complicated linkage to ensure anonymity. The applications and contact with each registry were handled by the Cancer Registry. Obtaining approval from these other registries took up to 18 additional months (august 2013), and the actual linkage a few more months. The Cancer Registry project database was linked to Statistics Norway (information on country of birth, county and municipality of residence, migration, causes of death, education, income and marital status), the Medical Birth Registry of Norway (births), the Norwegian Patient Registry (from 2008; information on treatment due to breast cancer), and the Norwegian Prescription Database (use of hormone therapy). Four project groups received such linked data from these other registries, and obtained these additional data during winter/spring 2014.

[43] Norwegian name: Datatilsynet

# 3 Project overview

## 3.1 Projects in the evaluation portfolio

The evaluation portfolio includes the following projects:

| Project no. | Sub-projects | Project Owner | Project manager |
|---|---|---|---|
| 189488 | Overdiagnosis | University of Oslo | Jan Mæhlen/ Nina Inger Farstad |
| 189494 | Screening costs<br>Modelling the treatment cost for breast cancer<br>Cost-effectiveness of mammography screening.<br>Coordinated with project number 189514 | University of Oslo | Ivar Sønbø Kristiansen |
| 189503 | Breast cancer mortality and overall mortality rates<br>Changes in age-specific and age-adjusted breast cancer stage distribution and histology<br>Compare trends in age-specific incidence and mortality from breast cancer in different European countries | Norwegian University of Science and Technology | Lars Vatten |
| 189504 | The Women's perspective | Norwegian University of Science and Technology | Siri Forsmo |
| 189505 | Evaluations of effects: breast cancer mortality, staging and histological grade, interval cancer, overdiagnosis<br>Other subjects to be analyzed: opportunistic screening, use of hormone therapy | University of Tromsø | Eiliv Lund |
| 189514 | Cost-effectiveness of breast cancer screening in Norway.<br>Coordinated with project number 189494 | Erasmus MC Dept. of Public Health | Harry de Koning |
| 189520 | Overdiagnosis | Barts and the London School of Medicine Wolfson Institute of Preventive Medicine | Stephen Duffy |

An overview of the publications and projects reports from each project group are summarized in Appendix III.

## 3.2 Identification of projects outside the evaluation

A literature search was conducted to identify relevant studies outside the evaluation portfolio on the topics addressed by the research groups in the above mentioned projects (mortality, overdiagnosis, interval cancer, stage distribution, the participants' experiences, costs and cost-effectiveness). Stage distribution was later excluded, since there were no publications from the research groups in the evaluation portfolio addressing stage distribution.
The following criteria for inclusion were set in advance:

- Date of publication in 2008 or later
- Peer-reviewed publications (i.e. letters to the editor and editorials were excluded)
- Presentation of original analyses and estimates based on data from the Norwegian Breast Cancer Screening Program or other original data from Norway
- For studies on the same topic and population by the same author, only the most recent study with the longest follow-up was included.

### 3.2.1 Search strategy

PubMed/Medline was search using the following search terms:

> Norw* AND (breast OR mammary) AND (screening OR "early detection" OR "early diagnosis")

A total of 592 publications from 2008 and later were identified until February 2015, including publications from the evaluation portfolio. Titles and abstracts were screened, and full-text publications were retrieved for all abstracts that contained information on any of the above mentioned topics. Other combinations of search terms such as mammog*, Norw*, breast, cancer, identified fewer studies. In addition, the Cancer Registry of Norway provided a list of all researchers that had received data on breast cancer from May 2007 to January 2013 and PubMed/Medline was searched by author name for these researchers. Searches by author names were ended in November 2013. Reference lists of the identified studies were screened for other relevant publications.

A total of 29 original publications outside the evaluation portfolio were included. Six of these were subsequently excluded since stage distribution was not included as a topic of the evaluation. The remaining 23 publications are summarized in chapter 5, together with studies from the evaluation portfolio.

# 4 Methodological considerations

## 4.1 Purpose of the evaluation and exposure of interest: intention to screen

The purpose of the evaluation is to assess the effectiveness[44] of the NBCSP at the population level as it evolved in real life, and not the efficacy[45] of mammography screening per se in the ideal trial situation. Estimating the effectiveness of the NBCSP may be compared to an intention-to-treat analysis in a randomized controlled trial, with receiving an invitation for mammography screening as the exposure of interest. The effectiveness of the NBCSP will depend on both the efficacy of the mammography screening program offered and the attendance in the program.

Comparing screening attendance as opposed to non-attendance in an invited population would introduce systematic error[46] through self-selection. The term self-selection is used to describe the fact that women who attend screening differ from those who do not attend in aspects that are related to their risk of breast cancer and/or their risk of dying from breast cancer. It has been reported that attendance varies according to age at invitation [83] and county [84], but knowledge on other determinants of NBCSP attendance is limited. Examples of factors known to influence breast cancer risk and/or prognosis include the use of hormone therapy [37], educational level [85, 86] and family history of breast cancer [87]. As described in section 1.1.5, women who use hormone therapy are recommended to attend NBCSP according to clinical guidelines [45]. Women with familial risk of breast cancer are recommended to attend NBCSP when they are no longer offered the tailored and age-limited breast cancer screening [11] (see section 1.3.2). Educational level is associated with use of health care services in general [88, 89].

A comparison of attending and non-attending women is therefore not an unbiased method to estimate the efficacy of the screening program, and we consider invitation status as the primary exposure of interest in this evaluation. The effectiveness of the program is therefore assessed in an intention-to-screen perspective.

## 4.2 Benefits of screening: effectiveness

In section 1.2, we described that mammography screening advances the time of breast cancer diagnosis with an inherently unobservable amount of time (lead time). Earlier detection of breast cancer will increase the time from breast cancer diagnosis to death from breast cancer, even if earlier detection would not postpone death. In consequence, statistical methods that use time from diagnosis to death in breast cancer, such as survival

[44] *Effectiveness* is the effect of implementing screening as a population-based program, i.e. the effect of inviting women for mammography screening.

[45] By *efficacy* we refer to the effect of screening in woman attending screening. Efficacy should preferably be investigated in an ideal randomized controlled trial with very high attendance after invitation.

[46] *Systematic error* refers to any distortion of the results away from the true estimate apart from random variation. Typical sources of systematic error include confounding, selection bias and information bias.

from breast cancer, cannot be used as an outcome measure when evaluating mammography screening effectiveness.

To avoid bias from lead time, breast cancer mortality is the most commonly used outcome measure in studies of mammography screening effectiveness. Invitation for mammography screening may prevent deaths from breast cancer only when breast cancer is diagnosed after screening invitation. To assess the effectiveness of mammography screening, calculation of breast cancer mortality should be restricted to deaths from breast cancer diagnosed after screening invitation (incidence-based mortality). This approach would correspond to the exclusion of women with known breast cancer at baseline in a randomized trial.

Although studies of incidence-based mortality are not affected by lead time bias in survival times after diagnosis, accounting for the effects of lead time on breast cancer incidence rates during the study period is still required. Since a large proportion of deaths from breast cancer diagnosed among women 50-69 years occur after 69 years, the full effect on breast cancer mortality can only be measured if women are followed for a period after they leave the screening program at age 69 years. However, due to earlier diagnosis, incidence rates will be higher among women offered screening compared to women not offered screening. Lead time bias in incidence-based mortality may arise if women are followed for death from breast cancer beyond the inclusion period of incident breast cancer diagnoses, since more breast cancer diagnoses and thus more deaths from breast cancer could be counted among women offered screening. As discussed by Njor et al [90], such bias may be avoided by ensuring that the accrual period[47] for breast cancer diagnosis and the follow-up period[48] for breast cancer deaths are equal. When follow-up extends beyond the screening age range, this means that deaths from breast cancers detected after the screening age range should also be included. In summary, to avoid bias from lead time, studies of incidence-based mortality should follow women invited for screening beyond the screening age range and include deaths from all breast cancers diagnosed during the entire follow-up period.

When the decision to offer a nation-wide mammography screening program was made, a relative reduction in breast cancer mortality of 30% was the primary aim. We are not familiar with any further specification of this aim in terms of background breast cancer mortality or in which age groups the reductions were expected. An effective program would be expected to reduce breast cancer mortality not only during the screening age range, but also a certain period after women have left the program. The evaluation age range could therefore be argued to extend beyond the upper limit of the screening age range. We also consider that a reduction in breast cancer mortality due to NBCSP implementation should be evaluated compared to the expected breast cancer mortality during the same time period in the absence of mammography screening. The gradual implementation of the program could provide opportunities for such a comparison, given that the observation period would be sufficiently long.

[47] Accrual period is the time period during which incident breast cancer diagnoses are included.
[48] Follow-up period is the time period during which deaths from incident breast cancers are counted.

## 4.3 Harms of screening

### 4.3.1 Overdiagnosis

In section 1.2, overdiagnosis due to mammography screening was defined as a breast cancer that would not be detected during the woman's lifetime in the absence of screening. Overdiagnosis is currently considered as the most important harm associated with mammography screening [14, 56]. Harms from overdiagnosis are closely linked to our lack of ability to identify overdiagnosed cancers at the individual level, resulting in unnecessary treatment and psychological distress for the individual women, and increased costs and work load in the health care systems. It is therefore important to quantify the extent of overdiagnosis associated with screening.

The extent of overdiagnosis in modern population-based programs is a topic of intense debate and there is considerable variation in methods used to quantify overdiagnosis [91, 92]. A certain amount of overdiagnosis is inevitable in a screening program that succeeds in advancing the time of diagnosis to a preclinical stage. Conversely, since only screening-detected breast cancer can be overdiagnosed due to the screening program, the number of overdiagnoses cannot exceed the number of screening-detected cancers.

During the screening period (50-69 years), incidence rates of breast cancer will be increased compared to a situation without screening. This increase will result from a mixture of earlier diagnoses of cancer that would otherwise be detected at later ages (lead time effect) and of tumors that would never be detected without screening (overdiagnosis). At later ages (70 years and older), incidence rates will be lower in a group offered screening compared to a situation without screening. This post-screening drop will be due to earlier detection (lead time effect) [92]. To estimate overdiagnosis, this lead time effect must be taken into account. Ideally, overdiagnosis should be estimated through comparison of cumulative risk of breast cancer in a large randomized clinical trial with high attendance, lifelong follow-up and no screening offered to the control group [14]. Any excess of breast cancers in the group invited for screening in such a study could be regarded as cancers overdiagnosed due to screening.

In observational studies, the following main approaches have been used to account for lead time [91]:

1) Comparison of the observed excess incidence during screening and the compensatory drop after screening to the expected incidence in the absence of screening (often termed 'excess-incidence approach')
2) Statistical modeling of lead time based on the excess incidence during screening (often termed 'lead time approach').

The validity of both approaches depends on the estimation of a usually unobservable incidence in the absence of screening [91]. In addition, the excess-incidence approach requires a long follow-up period for women who have previously been invited to screening in order to estimate the compensatory drop. In the lead time approach, unverifiable assumptions about the distribution of lead time are required. Overdiagnosed tumors would be expected to have longer lead times than other tumors, since they by definition must have a lead time longer than the woman's remaining lifetime. The previously described breast cancer heterogeneity also implies that one, common distribution of lead time for all breast cancers is an oversimplification.

Once the number of overdiagnosed cases (i.e. the lifetime excess number of cases in the invited group compared to the control group), has been estimated, there are several

possible measures (denominators) to express the extent of overdiagnosis. This has been illustrated by Marmot et al [14] and was also described in section 1.5.1. Overdiagnosis can be expressed as a proportion of breast cancers detected in a situation without screening among women followed from the start of screening and throughout life (i.e. the control group, termed method A). This measure indicates the load of overdiagnosis for the population compared to no screening. When expressed as a percentage of cancers detected after lifelong follow-up of a population invited for screening (method B), a measure of the population load of overdiagnosis in the current situation (i.e. with screening) is obtained. Overdiagnosis can also be expressed as a percentage of cancers detected while women are in the program (method C) and as a percentage of screening-detected breast cancers (method D). The latter two measures were considered by Marmot et al as helpful in decision-making for women invited to screening, since they express the risk that a cancer detected at screening or while receiving invitations for screening should be an overdiagnosed tumor.

The choice of denominator will affect the estimate even if the absolute number of overdiagnosed cases is similar. This may be illustrated by the following *hypothetical* example: Among 10 000 women invited for screening at age 50 years and followed throughout their remaining lifetime, 800 women were diagnosed with breast cancer. A total of 600 of these diagnoses were made during the screening age range (50-69), of which 400 were detected at screening. In a situation with no screening, there would be 650 breast cancer diagnoses from age 50 years and throughout the remaining lifetime for the same 10 000 women. The absolute number of overdiagnoses would then be 150 (800-650=150). Using method A above, there would be 23% overdiagnoses (150/650=0.23 or 23%), whereas using method B there would be 19% overdiagnosis (150/800=0.19 or 19%). Method C and D would give higher percentages due to smaller denominators, with method C 150/600=0.25 or 25% and method D 150/400=0.38 or 38%.

There is no consensus on which denominator should be chosen. As can be seen from the example above, expressing overdiagnosis as a proportion of breast cancer in a population offered screening will result in a lower estimate than choosing a population not offered screening as the denominator. Similarly, choosing breast cancer detected during the screening age range as denominator will give a higher estimate than breast cancer detected during lifelong follow-up.

Both screen-detected invasive tumors and screen-detected DCIS may represent overdiagnoses with subsequent unnecessary treatment. Since treatment of DCIS may prevent some cases of invasive breast cancer, the lifetime occurrence of DCIS and invasive breast cancer in a population offered screening should be viewed as dependent [54]. Estimates of overdiagnosis should therefore include both invasive tumors and DCIS combined.

### 4.3.2 Interval cancer

Surveillance of interval cancer rates in a screening program is also important to evaluate the performance of the program. Interval cancers that were missed at screening (false negative) provide information on the sensitivity of the mammography examination. Interval cancers following true negative mammograms (true interval cancers) can be used to evaluate the duration of the screening interval. The distinction between these two groups of interval cancer requires review of the previous screening mammogram of women with an interval cancer.

A negative mammogram may provide a sense of security for women participating in screening. For women who are subsequently diagnosed with interval cancer, the negative mammogram would represent a false reassurance. The key concern is that the reassurance provided by a negative mammogram or the prospect of an upcoming screening invitation might lead to diagnostic delay for women that develop symptoms of breast cancer between screening rounds. Diagnostic delay could lead to poorer outcome.

Interval cancers tend to have a less favorable distribution of tumor characteristics at the time of diagnosis compared to cancers detected at screening [62]. This could result both from diagnostic delay and from more rapid growth patterns of true interval cancers. In comparisons between interval cancers and other clinically detected breast cancers, it should also be kept in mind that women with interval cancers are screening attenders and that factors contributing to screening attendance may influence prognosis, tumor characteristics, and breast cancer risk (self-selection, described in section 4.1).

### 4.3.3 Recall examinations

Women with positive screening mammograms are recalled for further diagnostic tests. In 2012-2013, 11 400 (2.8%) of the women attending NBCSP were recalled [63]. Among recalled women, only a proportion will get a cancer diagnosis (18.9% in 2012-2013). The remainder will have false positive mammograms. These women will experience mental and physical distress caused by the possibility of having cancer and by the additional test, which would not have occurred in the absence of screening. Quantification of the probabilities of false positive results and knowledge on how a false positive episode influences the quality of life is therefore important. Recall examinations also add to the costs of screening and work load in the health care systems.

From a public health view, the rates of recall examinations within a single screening round may be useful when evaluating costs and work load in the screening program. In decision-making for an individual woman, the cumulative probability of a false positive result during the screening period may be more relevant [54].

It should be noted that the number of false positive recalls and the number of false negative mammograms are dependent, and that reducing the number of recall examinations can only be obtained at the expense of more interval cancers (and vice versa) [54].

### 4.3.4 Women's perspective on screening particiption and recall examinations

When individuals consult health care services due to health related problem, they have decided to seek help based on their complaints, knowledge of and experiences with the health care services. The reflections on whether to attend a national screening program are quite different since the invited women are healthy with no signs of breast cancer disease. The information given in the invitation must be trustworthy and understandable so the women may decide on whether to attend or not. For a national screening program the attendance rate highly influences the effectiveness of the program. It is therefore important to reveal the women's premises for their decisions on attending or not. Further, it is important when evaluating the benefits and harms of the program to know women's experience of having a normal mammogram, to be recalled due to a false positive or a true positive mammogram, and eventually the experience of an interval cancer, either a

false negative or a true interval breast cancer. Ideally, information from all these groups is desirable to make a complete evaluation of the women's perspectives.

The two major designs for collecting such data are quantitative surveys, mainly based on registry data and questionnaires, and qualitative studies, interview or focus group based. Surveys may identify individual, socioeconomic, and structural factors predicting participation [93-95]. There are some specific questionnaires developed for e.g. anxiety among recalled women [96]. In general, questionnaires have limitations such as response rates, lack of validation, and predefined response categories, and in retrospective studies of experiences with the screening program, recall bias.

Qualitative studies, either individual or focus group interviews, have the advantage to collect not preconceived information, and to explore the reasons for their point of views. The challenges are to achieve information from all groups ascribed over, to continue interviewing until information saturation, and to use transparent and verifiable methods of interpretation of
the collected information.

## 4.4 Cost-effectiveness

The NBCSP is evaluated from the perspective of society. An accurate assessment of the cost-effectiveness of the NBCSP requires that all costs and effects associated with the program are identified.

### 4.4.1 Costs

The NBCSP generates different types of costs, some of which are costs associated with administrating the program as such and communication with participants (e.g. information, invitations, reminders and recalls). Moreover, there are costs of running the mammography-units and -buses, costs of mammography examination, costs of diagnostic evaluation of abnormal findings on mammograms, storage and management of screening results, etc. Participation in the NBCSP also implies costs of travelling, and potentially also indirect costs of productivity loss (due to absence from work).

The NBCSP may alter the health care costs associated with breast cancer treatment relatively to a situation without a screening program. Earlier detection of malignant tumors may result in less aggressive and possibly less costly treatment, while overdiagnosed tumors may increase health care costs compared to a situation without screening. The NBCSP may indirectly also affect the use of resources in other sectors, for instance primary health care and the Norwegian Labour and Welfare Administration[49] (e.g. sickness benefits).

### 4.4.2 Benefits and harms

The main objective of the NBCSP is to reduce the number of deaths and the number of life years lost from breast cancer. Hence the effects of the program may be measured as the number of prevented deaths, or the number of life-years saved. Often, adjustment for impaired quality of life associated with cancer treatment and progression of disease, is

[49] Norwegian name: Arbeids- og velferdsforvaltninga (NAV)

taken into account. Quality adjusted life-years (QALYs) are a measure that combines the number of life-years saved and a measure of the quality of life during those gained years. QALYs thus capture benefits from both reduced morbidity and reduced mortality. There are considerable challenges in assessing quality of life. Earlier diagnosis (benefits) and overdiagnosis (harms) will influence the population estimate in opposite directions, and the balance between these would need to be considered.

Individuals participating in the NBCSP may place value on the information gained from undergoing screening. Negative findings (no tumor) provide reassurance, but the value of reassurance is difficult to monetize. On the other hand, positive findings (tumor suspicion) may generate concern.

### 4.4.3 Cost-effectiveness

The cost-effectiveness of NBCSP may be indicated by the ratio of its costs and a quantitative measure of its benefits. If the objective of the NBCSP is quantified in terms of number of lives saved from breast cancer mortality, the cost-efficiency ratio gives costs per life-year saved. The efficiency of the NBCSP may also be measured in terms of number of quality adjusted life-years saved. The cost-efficiency ratio then gives costs per quality adjusted life-year saved.

Information about all relevant costs and benefits of the NBCSP program is not easily available, and there are methodological challenges in assessing some of these costs and benefits (see section 4.1). Estimating the cost-effectiveness ratio consequently provides an estimate only of the program's cost-effectiveness. When interpreting the estimated cost-effectiveness ratio, one should be aware that in addition to the limitations inherent in the estimate itself, the evaluation of the program cannot be based on its cost-effectiveness ratio alone. In addition, alternative programs or health services that can be provided at an even lower cost-effectiveness ratio, if the resources now being spent on NBCSP were re-allocated, would need to be taken into account. Evaluation of alternative use of resources spent on NBCSP is beyond the scope of this report.

In Norway, there is no official threshold value of a QALY, nor is there an official threshold value at which a program is considered cost-efficient or not [97]. The Norwegian Directorate of Health indicates that one QALY has a value of NOK 1 million (2012), based on an economic value of a statistical life equal to NOK 30 million.

## 4.5 Study design

The optimal study design to evaluate the effectiveness of implementing a mammography screening program would be a large, randomized controlled trial with a long follow-up period and no mammography screening offered to the control group during the study period. In contrast, the NBCSP was implemented as a pilot project in four non-randomly selected counties in 1995-96, with gradual expansion of the program to reach nationwide coverage by the end of 2005. With a fully implemented population-based program, evaluation through randomized trials is not realistic. As a result, only observational studies can be conducted to specifically assess benefits and harms in the NBCSP. In testing the effectiveness of a medical intervention, observational studies are considered to

provide a weaker level of evidence than do randomized controlled trials, due to a higher risk of bias[50] and confounding[51].

Among observational studies, analytical studies such as cohort studies with individual information are better suited for causal inference than ecologic studies. In cohort studies, individuals with different levels of exposure to the factor of interest are followed over time and compared with respect to occurrence of the study outcome of interest. In the context of mammography screening, women invited for screening could be compared to women not invited with respect to the risk of death from breast cancer. An essential difference between a cohort study and a randomized controlled trial is that in a cohort study, the exposure status is not randomly allocated, but based on a decision by the participant or her health care providers. In consequence, systematic differences between the groups apart from the difference in exposure may distort the results.

Ecologic studies[52] are often termed aggregate or descriptive studies and are characterized by comparison of groups or populations rather than individuals. Ecologic studies may be used to study trends in disease occurrence over time or geographical differences. In an ecologic study, some or all of the study factors are measured at the group or population level. The population's exposure or outcome levels may not reflect the association between an individual's exposure status and her outcome risk [98]. The failure of an ecologic association to reflect the causal effect at the individual level is often termed ecologic bias. In the context of mammography screening, a decline in breast cancer mortality for the population following the implementation of a population-based screening program need not reflect an effect of screening for the invited individuals, but could also reflect a general focus on cancer management leading to both screening implementation and improved treatment. Conversely, lack of reduction in breast cancer mortality at a population level following screening implementation need not reflect program ineffectiveness. Instead, the finding could be due to failure of aggregate data to separate deaths from breast detected within the program form those occurring outside the program, or to concurrent factors increasing the risk of death from breast cancer in the population.

### 4.5.1 Individual data

As previously described, the NBCSP was implemented gradually across the country and invitations are distributed evenly across the entire two-year period of each screening round. Invitation dates are organized according to municipalities and six-month periods. In consequence, within each municipality, the date of invitation can vary by six months, and within each county, the date of first invitation can vary by two years. Invitations are organized according to birth cohorts, with the implication that some women received their first invitation at 48-49 years and some receive their last invitation at age 70-71 years. To avoid misclassification of exposure status (invitation for screening), individual data on invitation date are essential.

[50] Bias refers to a distortion of the effect estimate away from the true estimate due to errors in the design or conduct of the study

[51] *Confounding* refers to a distortion of the effect estimate due to mixing of extraneous effects and the effect under study

[52] *Ecologic studies* are studies in which two or more of the study factors are measured at the group or population level rather than at the individual level.

### 4.5.2 Duration of follow-up period

The previously described heterogeneity of breast cancer, with substantial variations in tumor characteristics and growth, also within the group of patients who eventually die from the disease, underscores the necessity of a long observation period when studying changes in breast cancer mortality. By definition, a reliable assessment of overdiagnosis would also require a very long observation time in order to determine the proportion of tumors that would or would not be detected during the remaining lifetime.

## 4.6 Potential challenges in observational studies of the Norwegian Breast Cancer Screening Program

To ensure accuracy in estimation of benefits and harms of screening, both random and systematic errors should be reduced as far as possible.

### 4.6.1 Information bias and misclassification of study factors

Information bias may result from errors in measurement of the study factors. Errors that depend on the value of other study factors are known as differential misclassification, and may bias the results in any direction. Errors that do not depend on the values of other factors are termed non-differential misclassification, and will most often lead to an underestimation of the association since it increases similarity between the compared groups [99].

*Information in the Norwegian Breast Cancer Screening Program database*
The NBCSP database was established in parallel with the implementation of the program and contains information on all program activity such as invitation, attendance, interpretation of screening mammograms, recall examinations and information from questionnaires [70]. The Cancer Registry of Norway is responsible for quality assurance of the database and linkage with the incidence register at the Cancer Registry of Norway is part of the quality assurance procedure. The extensive quality assurance work that was initiated at the start of this evaluation has been described in detail in section 2.5. A description of the project database is provided in Appendix II.

*Information on causes of death*
Accurate classification of death from breast cancer is essential for a valid estimate of NBCSP effectiveness. Medical doctors have reported causes of death to the Cause of Death Registry since 1951. All deaths are reported on standardized death certificate forms, and reporting is mandatory [100]. Causes of death are registered according to ICD-codes, and both the underlying and other causes of death are reported. To ensure completeness, information in the Cause of Death Registry is regularly cross-checked with information on vital status from the Central Population Register, and information from the Cancer Registry of Norway and hospital records of postmortem examinations [101]. Revision studies of the reported causes of death indicate that registration of deaths from cancer has been reliable over time [102, 103]. Since cause of death is registered independently from screening invitation and results of screening, any misclassification

would most likely be non-differential and lead to underestimation of screening effectiveness.

*Breast cancer screening outside the NBCSP*
As described in section 1.3.2, there has been a gradual increase in the availability and use of mammography in Norway since the 1980s, but with large geographical variations. Although a proportion of the mammography examinations conducted outside the NBCSP are diagnostic and should not be considered as a substitute for program screening, it is seems clear that the extent of non-organized screening prior to and in parallel with NBCSP implementation may not be ignored in studies of effectiveness and overdiagnosis. Assessing the impact of misclassification of exposure status due to non-program screening on breast cancer incidence and mortality is complicated by the fact that the extent, sensitivity and effectiveness of such screening is not known. We also have little information on the characteristics of users of non-program screening.

The direction of the bias from non-program screening would depend on the study design and on the distribution of non-program screening in different age groups and according to invitation status. However, the overall expectation when introducing a screening program in an already partially screened population would be that both effectiveness and overdiagnosis would be underestimated. Continued use of non-program screening among invited women, both before and after reaching the upper age limit of the screening program, could lead to overestimation of effectiveness and overdiagnosis. In studies of incidence-based mortality, non-program screening before entering the study would to some extent shift cancer diagnosis out of the study period and thereby reduce incidence-based mortality.

### 4.6.2 Selection bias

Common sources of selection bias are situations where the relation between exposure and the outcome of interest differs between the subjects who are part of the study and the subjects who in theory could have been part of the study, i.e. the study participants are not representative of the source population [99]. Selection bias may arise for example in comparisons of women attending screening with all women invited for screening or comparing women in specific regions or counties to women in the entire country.

### 4.6.3 Challenges in the choice of comparison groups

One of the greatest challenges when designing an observational study to compare benefits or harms in the NBCSP is to select a suitable unexposed comparison group. If the exposed group consists of women invited for screening, the unexposed group should ideally reflect the situation for the group of invited women if screening had not been implemented. The gradual implementation necessitates a comparison between different regional characteristics and time periods with different treatments and distributions of risk factors for breast cancer.

*Comparison between women in different regions or counties*
The four pilot counties (Oslo, Akershus, Rogaland and Hordaland) represented 40% of the Norwegian population in 1996, but should not be considered to represent a 40% random sample of the population. There are important demographic and socioeconomic

differences between the population in these counties and the rest of the country. The four pilot counties are the most urban, centralized parts of Norway, with a higher population density, educational level and life expectancy than most other counties [50]. Before the implementation of NBCSP, the pilot counties had higher breast cancer incidence rates in the screening age range [104] and lower breast cancer mortality among women 50 years or older [105] than other counties in Norway. The higher incidence rates may at least in part result from a less favorable distribution of breast cancer risk factors, such as a higher consumption of hormone therapy [34] and a higher age at first birth [106]. Differences in mortality could reflect differences in breast cancer management and use of non-program screening. The differences between the pilot counties and the remaining counties could result in biased estimates if using women in the non-pilot counties as an unexposed comparison group.

*Comparison between women in different time periods*

Another approach to select a non-invited comparison group could be to study changes over time within counties. However, this would also require assumptions about the development over time in each county in the absence of NBCSP implementation. Such assumptions would present special challenges in the analyses, since several factors expected to influence breast cancer incidence rates and/or breast cancer mortality changed substantially in parallel with NBCSP implementation.

Table 2. Incidence rates of invasive breast cancer in Norwegian counties for women aged 50-69 years in 1991-1995, from highest to lowest

| County | Incidence rate per 100 000 person-years | Start of NBCSP implementation |
|---|---|---|
| Oslo | 238 | January 1996 |
| Sør-Trøndelag | 188 | September 2001 |
| Akershus | 187 | March 1996 |
| Hordaland | 185 | January 1996 |
| Rogaland | 185 | November 1995 |
| Vest-Agder | 179 | November 1999 |
| Buskerud | 177 | August 2001 |
| Møre og Romsdal | 176 | March 2002 |
| Hedmark | 175 | August 2003 |
| Nord-Trøndelag | 174 | September 2001 |
| Oppland | 174 | January 2002 |
| Østfold | 163 | April 2001 |
| Aust-Agder | 162 | November 1999 |
| Nordland | 157 | May 2001 |
| Vestfold | 156 | February 2004 |
| Telemark | 155 | August 1999 |
| Sogn og Fjordane | 142 | February 2003 |
| Troms | 136 | April 2000 |
| Finnmark | 130 | April 2000 |
| All counties combined | 179 | |

*Improved treatment and management of breast cancer*
Over the last decades, there have been important developments in breast cancer treatment, as described in chapter 1.1.3, that could be expected to influence breast cancer mortality. Furthermore, the reorganization of breast cancer management in terms of centralization, multidisciplinary teams and devoted breast cancer centers that took place prior to screening implementation in each county, may also have improved prognosis from breast cancer. Comparing breast cancer mortality for women in different time periods according to invitation status should attempt to take into account concurrent changes in breast cancer mortality in the absence of screening.

*Changes in use of menopausal hormone therapy*
The dramatic increase in use of hormone therapy in parallel with screening implementation and the subsequent decline (see section 1.1.5) would be expected to increase the incidence rates of breast cancer during the period. To estimate the incidence in the absence of screening, which is required in studies of overdiagnosis, changes in hormone therapy use should be taken into account. Unfortunately, nation-wide individual data on hormone therapy prescriptions have not been available throughout the relevant study period since the Prescription Database contains individual level information only from 2004.

In addition to higher incidence rates of breast cancer, women who use hormone therapy are also at increased risk of dying from breast cancer, partly due to the higher risk of breast cancer in itself, and partly due to less favorable tumor characteristics [40, 41]. Changes in hormone therapy use may therefore also be expected to influence the mortality from breast cancer in the absence of screening.

Use of hormone therapy is associated with high mammographic density [107]. High mammographic density is associated with reduced mammographic sensitivity [108]. In consequence, women who use hormone therapy are at higher risk of having false negative screening mammograms and thus increased risk of interval cancer [38]. Women who use hormone therapy are also at higher risk of false positive mammograms and recall examinations [109].

As described in section 1.1.5, awareness of an increased risk of breast cancer associated with hormone therapy has also resulted in treatment guidelines emphasizing the need for regular examination of the breasts [43, 45]. Thus if hormone therapy use and patterns of change use are associated with exposure to screening, this may influence estimates of screening effectiveness, overdiagnosis, rates of interval cancer and false positive mammograms.

### 4.6.4 Precision

Random error in an estimated association is a measure of the expected fluctuations around the true population value of the association due to chance, reflecting that some of the underlying processes are stochastic. Random variation may be expressed by the variance of the estimate. Statistical precision may be thought of as the opposite of random error. Measures of statistical precision include standard error and confidence intervals. Precision increases with increasing number of study participants and increasing number of events (such as deaths from breast cancer) and also depends on the ratio of exposed to non-exposed (invited and non-invited) participants [110].

# 5 Results

This chapter provides a summary of the results for each topic in the evaluation – breast cancer mortality (section 5.1), overdiagnosis (section 5.2), interval cancer (section 5.3), risk of false-positive mammograms and the experiences of women who participate in screening without being diagnosed with breast cancer (section 5.4) and costs of screening and cost-effectiveness (section 5.5). Within each topic, studies that are part of the evaluation portfolio are described separately from those that are not part of the evaluation portfolio.

For each study, a summary of the methods and results are presented, followed by our evaluation of the study in terms of strengths and weaknesses. These may or may not agree with the authors' own assessments. In the end of each chapter, a summary across all the studies are presented together with our views on what are the most reliable results. Studies that were submitted to the Research Council of Norway, but were not part of the original contract with the research groups, will be commented more briefly.

## 5.1 Breast cancer mortality

### 5.1.1 Studies in the evaluation

**Breast cancer mortality in Norway after the introduction of mammography screening.**
Olsen AH, Lynge E, Njor SH, Kumle M, Waaseth M, Braaten T, Lund E.
Int J Cancer. 2013 Jan 1;132(1):208-14.

*Summary of methods and results*
The authors compared incidence-based breast cancer mortality for invited women in the 4 pilot counties (Oslo, Akershus, Rogaland and Hordaland, termed study group) and not yet invited women in the counties included from 2002 and later (Oppland, Hedmark, Vestfold, Møre og Romsdal and Sogn og Fjordane, termed regional control group). First invitation date was measured at the municipality level, while the remaining data were at an individual level.

Two different approaches were used in the comparison of incidence-based mortality, termed the 'Follow-up model' and the 'Evaluation model', as described by Nyström et al [111]. In the 'Follow-up model', the incidence period was from the first invitation in 1996 until the end of 2001 for women aged 50-64 years when screening started (maximum 6 years), and from the first invitation to the end of 2008 for women aged 65-69 years when screening started (maximum 13 years). The follow-up period for death from incident breast cancer was equal to the incidence period in this model.

In the 'Evaluation model', the incidence period for all women was restricted to 1996-2001, i.e. only incidence during the screening age range was counted. Follow-up was similar to that in the 'Follow-up model'.

To account for the underlying temporal decline in breast cancer mortality in the absence of screening, the rate within each region was first compared to the rate in a

preceding period of similar duration as the study period (1990-1996 for women aged 50-64 years and 1990-2002 for women aged 65-69, termed historical control groups) and this ratio was subsequently compared between the screening and non-screening counties. The rates were compared using Poisson regression and adjusted for (current) age in 5-year groups.

When counting all breast cancer deaths from disease diagnosed after the first invitation date in each municipality and throughout follow-up (the 'Follow-up model') the authors found a rate ratio of 0.93 (95% CI 0.77-1.12). When counting only breast cancer deaths from disease diagnosed during the screening age range, 50-69 years, but still including deaths throughout follow-up (the 'Evaluation model') the rate ratio was 0.89 (95% CI 0.71-1.12).

*Characterization, strengths and limitations*

This study used a design developed and implemented in Denmark by Olsen and Lynge, where the effect measure is a double ratio of rates:

(Study / concurrent control) / (historical study / historical control)

The design has well-established parallels in econometrics (the 'dif-dif' design [112]). The Danish documentation [111, 113, 114] emphasizes that the design enables the analysis of screening *versus* non-screening controlled for historical trends, but that this requires the assumption of no interaction between historical development and area.

The study design allows for a detailed account of underlying person-years and a near complete enumeration of breast cancer deaths as well as at least a partial control for underlying time trends in breast cancer mortality.

The present implementation is inconclusive regarding the choice between Nyström's 'Follow-up model' and 'Evaluation model'. Njor et al [90] discussed the possible lead time bias of the 'Evaluation model'. The 'Evaluation model' produces the strongest effect measured on the relative scale chosen. Nyström's evaluation model was proposed in a different context from the Norwegian situation: Nyström et al compared long-time follow-up in a randomized trial between screened and non-screened controls, where the latter had a prevalence screen at the end of the first trial period. As explained by Njor et al [90] analyses where the accrual period and the follow-up period (defined in chapter 4.2) do not coincide are prone to lead time bias. With earlier diagnosis due to screening, more women will have a diagnosis of breast cancer at 50-69 years in the study group than in the control groups, and will be at risk of death from breast cancer during follow-up. This will lead to underestimation of the effect of screening. Although the addition of breast cancer cases diagnosed after the age span of screening seems to be neutral in the comparison between screened and controls, this is only correct if the effect measure were on an additive scale – however here the effect was measured on a multiplicative scale, generating an estimate of lower effect if cases diagnosed late are included.

Since date of first invitation to screening is measured at the municipality level (first day of screening in each municipality) and not individually, some women who were diagnosed with breast cancer before they received their first invitation (i.e. clinically detected tumors) will be classified as detected after invitation to screening. This will lead to underestimation of the effect of screening.

Attention is given to the possibility that any screening effect has been diluted by frequent use of non-program mammography in the control groups. There is disagreement

about the ability of the available Norwegian data to assess the amount of non-program mammography and how much of this can be reasonably considered to act as replacement for screening and extensive use in the control groups could reduce the estimated effectiveness by the organized program. However, in studies of incidence-based mortality a shift of cancer detection towards years before the incidence period will deplete the population during the incidence period of some breast cancers and thus remove breast cancer deaths in such cases from the population.

Due to the short period between start of screening in the pilot counties and in the last counties, the contrast in exposure (screening invitation) between the groups will be no more than one to three screening rounds. In combination with a short follow-up for most of the participants, this may have led to underestimation of the effect of the screening program. Since the study was restricted to the first and last counties where screening was implemented, statistical power was limited.

As mentioned above, the model is valid under the assumption of no interaction between period and region, i.e. that the relative decline in mortality over the period should be similar in both regions in the absence of screening. The pilot counties had higher breast cancer incidence rates [104] and lower breast cancer mortality than the rest of the country before screening was implemented. In addition, breast diagnostic centers and multidisciplinary teams were established before screening could start in each county, and would be functional from different time points in the pilot counties and the remaining counties. This may imply that some of the observed risk reduction may be attributed to multidisciplinary teams since any mortality benefit from specialized centers will be greater in the pilot counties, where such centers were present during the whole follow-up. There could also be trend differences in mortality at the county level due to differences in risk factors. A risk factor that differed between counties during the study period was hormone therapy use, which was more prevalent in the pilot counties than the control counties [34]. Since a prescription of hormone therapy has been followed by a recommendation of mammography examinations, it might also have increased the use of mammography outside the program.

*Conclusions*

The effect of screening is estimated as an overall non-significant rate ratio of between 0.89 and 0.93, these estimates being based on two different concepts for interpreting the data, and there is uncertainty which model would be preferable. The double ratio design has allowed correction for historical trends, but necessitates an assumption of no interaction between county and time trend. There are systematic differences in socio-demographic composition and development of health services between screening and control counties, which may distort the comparison in directions that are hard to predict. A widespread non-program mammography use and lack of individual data on invitation date may dilute any screening effect. Duration of both the exposure period and the follow-up period may be too short to provide information on the long-term effectiveness of a fully implemented screening program. Finally, it is hard to distinguish any effect of screening from effects of the concomitant organization of multidisciplinary specialized teams.

**Mammography activity in Norway 1983 to 2008.**

Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, Lund E.
Acta Oncol. 2011 Oct;50(7):1062-7.

*Summary of methods and results*

This is a report of total mammographic activities. The aim of the study was to investigate the extent of all mammographic activity in Norway during the period 1983-2008, and to estimate the impact of that activity on the effectiveness of NBCSP. The authors used publicly available numbers and data from NOWAC to estimate the number of women who had opportunistic screening prior to NBCSP implementation. Public sources included reports from the Norwegian Radiation Protection Agency (NRPA), summary data from the NBCSP, a report from the Cancer Registry of Norway and a report from the University College in Oslo. The NOWAC sample included women who responded to a questionnaire with questions about mammography use in 1996, 1997-1998, and 2002. It is stated that of 121 683 invited women in screening relevant ages (this is not specified, but data are presented for ages 40-69), 70% responded to the mammography questions. However, data are presented for 94 211 women, which is 77% of 121 683. In 1996 and 1997/1998 women were asked about regular use of mammography, and could answer no, yes – every second year or more often, and yes – with an interval of more than two years. In 2002, the question was about ever/never use of mammography.

The total number of mammography examinations in Norway registered by the NRPA was 10 000 in 1983, 80 000 in 1988, 221 210 in 1993, and 349 057 in 2002. For 1983-1993, the number of women examined was estimated by dividing the number of examinations by two, giving 5000 women examined in 1983, 40 000 in 1988, and 110 605 in 1993. For 2002, an algorithm developed by Hofvind was used to estimate that 131 758 women were examined outside NBCSP. The annual number of examined women in 1996 was also estimated from the NOWAC data, and compared to the numbers from NRPA. In 1996, 25% of 11 819 NOWAC participants 40-69 years had regular mammography with no more than 2 years interval, and 18% had regular mammography with more than two years interval. With the assumption that more than two years interval can be regarded as every fourth year, the authors estimate that 119 000 women aged 40-69 years were examined in Norway in 1996. The authors conclude that this number is comparable to the 110 605 women examined according to NRPA in 1993, and that the different sources of data collectively indicate that at least 40% of Norwegian women had regular mammography prior to their first NBCSP invitation.

Only the NOWAC data could be used to examine mammography use separately for the pilot counties and the non-pilot counties. In 1996, 47% of NOWAC participants aged 50-69 used mammography regularly in the pilot counties, compared to 40% in the non-pilot counties. In 1997-98, the numbers were 73% and 47%, respectively. In 2002, 97.5% of NOWAC participants 50-69 years in the pilot counties reported ever use of mammography, compared to 87.4% of women in the remaining NBCSP counties (combined 92%), and 79.3% in the counties where screening had not yet been implemented. In the NBCSP, 64% of attending women reported ever use of mammography before their first NBCSP attendance.

The authors used these numbers to estimate that a true risk reduction by screening of 25%, would be observed in a dif-dif design with three unexposed groups as only 11% reduction with their estimated level of opportunistic screening before and in parallel with the NBCSP implementation. The assumptions in those calculations were: Breast cancer mortality rate of 71/100 000 person-years before screening and 68/100 000 person-years

during the screening period if screening had not been implemented; similar effects of program and opportunistic screening; 40% opportunistic screening in 1996 across the country; 92% screened in the NBCSP counties and 64% screened in the non-NBCSP counties after 1996.

*Characterization, strengths and limitations*

This is a primarily descriptive study with comparison of information on mammography use from several sources.

The study provides an overview on the available information on opportunistic screening in Norway. The data from NOWAC have not been presented elsewhere and adds new information to a field with scarce information.

The comparison of regular non-program mammography in the pilot counties and other counties in different years are based on small numbers from the NOWAC. The questions on mammography use in 1996 and 1997-98 had no clear response alternative for women who had used mammography, but not at a regular basis. Furthermore, the question in the 2002 questionnaire did not measure regular mammography use, and the answers to this question should not be directly compared to those from 1996 and 1997-98. No information is provided concerning the algorithm used to estimate the number of women examined in 2002 from the NRPA data.

The comparison between the estimated 119 000 number of women 50-69 years examined with mammography in 1996 and the 110 605 women examined in 1993 according to NRPA, does not take into account that the number provided by NRPA includes all age groups and also clinical mammography. This implies that the estimate from NOWAC data may be quite a lot higher than the NRPA data from 1993. On the other hand, some NOWAC participants in 1996 had probably attended the NBCSP before responding to the questionnaire.

In the calculations of the expected effect of NBCSP in the presence of extensive opportunistic screening, the reported proportion of women with ever use of mammography were used as a measure of the proportion screened (through NBCSP or opportunistic screening). It is not known how many of the 64% NBCSP attendants reporting ever use of mammography prior to NBCSP should be classified as opportunistically screened, since only the time since previous mammography has been reported [76]. In addition, non-program screening could influence incidence-based breast cancer mortality in opposite directions through different mechanisms, as discussed in the evaluation of the study by Olsen et al above.

*Conclusions*

The study documents widespread use of mammography screening outside the NBCSP, but how much of this should be considered as non-program screening and the expected effectiveness of such screening remains unclear. The use of mammography in any choice of control group from the appropriate age groups in Norway would be higher than has been the mammography use in the control groups in the historic randomized trials.

**Modern mammography screening and breast cancer mortality: population study.**
Weedon-Fekjær H, Romundstad PR, Vatten LJ.
BMJ. 2014 Jun 17;348:g3701.

*Summary of methods and results*

The authors compared breast cancer mortality among invited and not invited women in an open cohort consisting of all Norwegian women followed while they were aged 50-79 years during 1986-2009. Invitation date for screening, date of breast cancer diagnosis and date of death from breast cancer were measured at an individual level. Person-time was measured at an ecologic level for each combination of calendar year, birth cohort and county.

Poisson regression with adjustment for age, birth year, calendar year at death, and county was used to estimate the mortality from breast cancer for invited women and non-invited women. To estimate the breast cancer mortality from 1996 to 2009 attributable to breast cancers diagnosed after invitation (incidence-based mortality), the authors used a model offset to adjust for the expected proportion of deaths caused by breast cancer diagnosed during the time period since first invitation. To avoid bias from lead time, this expected proportion was estimated from the distribution of time from diagnosis to death from breast cancer among women diagnosed before screening invitation, assuming that the time from diagnosis to death from breast cancer would be equal in the absence of screening. The model offset was estimated separately for women 50-59, 60-69 and 70-79 years at death and based on pre-invitation diagnoses from two different periods; 1990-1994 and 1996-2009, with very similar results.

Calculations based on simulated data confirmed the validity of the approach. Several sensitivity analyses were conducted to assess the impact of different assumptions, such as different ways of including covariates, varying the effect of screening invitation by calendar year, time since first invitation and time since last invitation, as well as expansion of the study age group to 40-85 years. Numbers needed to invite for screening to prevent one death from breast cancer was calculated according to the Stanford CISNET model [115, 116] using the national breast cancer mortality in 2009 and the estimated relative mortality reduction as the basis for the calculation.

Breast cancer mortality after invitation to screening was 28% lower than for women who were not invited (MRR 0.72, 95% CI 0.64 to 0.79). A lower mortality was observed also after screening invitations had ended, although less pronounced (for women 75-79 years MRR was 0.79, 95% CI 0.57 to 1.01). Inclusion of covariates did not change the estimates substantially. Most sensitivity analyses gave the same result as the primary analysis. Numbers needed to invite to prevent one death from breast cancer among women 50-89 years was estimated to 368 (95% CI 266 to 508).

*Characterization, strengths and limitations*

This study is based on an original epidemiological-statistical approach developed by the authors and documented in web appendices to this BMJ article. The approach – a Poisson regression model with non-screening breast cancer mortality as a shared underlying latent variable – allows inclusion of the individual experience of all women in the relevant age groups over a period (1986-2009), thus including a significant period before screening started.

The developed method may be considered as an original modern version of the classical technique of indirect standardization based on comparing mortality in a study group to that expected if the mortality was as in the standard (here: non-screening) group.

The development is supplemented with tests on simulated data (confirming that the postulated effects may be recovered by this approach) as well as careful and wide-ranging sensitivity analyses to assess the various model choices, documenting considerable robustness of the approach. The authors developed a detailed protocol before they had access to the data and deposited this protocol with the Norwegian Research Council to ensure that the modeling was not unduly influenced by fishing expeditions in the data. The study was conducted according to the study protocol.

Misclassification of invitation date should be minimal, since this was measured at an individual level. The extent of exposure misclassification due to non-program mammography is unknown, and the potential impact on the results is difficult to assess. As discussed in the evaluation of the study by Olsen et al, non-program screening in women not yet invited may influence the estimated association in different ways: Preventing some deaths from breast cancer in women whose cancer would have been clinically detected *before* their first invitation may lead to underestimation of the program's effectiveness. Earlier detection by opportunistic screening will move some breast cancer diagnoses to the non-invited group that would not be clinically detected until *after* the first invitation, which may lead to overestimation of the effectiveness. Non-program screening after the first invitation may also lead to overestimation of the program's effect.

Improved management resulting from the establishment of breast diagnostic centers and multidisciplinary teams would have a larger influence on breast cancer mortality in the invited group, where all breast cancer patients would receive this treatment, compared to the non-invited group, where only those diagnosed after screening implementation in their county but before they themselves were invited, would benefit from this.

*Conclusions*

This original contribution to the issue of assessing the effect of invitation to screening on incidence-based mortality yields an estimate of a relative mortality of 0.72 (95% CI: 0.64-0.79). Numbers needed to invite to prevent one death from breast cancer among women 50-89 years was estimated to 368 (95% CI 266-508).The analytical treatment is careful and wide-ranging sensitivity analyses indicate that the approach is robust to realistic deviations from model assumptions. As all other approaches, this cannot separate the effect of invitation to screening neither from an effect of the concomitantly established breast diagnostic centers, nor from an effect of opportunistic screening (which in principle could work both ways).

**Research-based evaluation of the Norwegian mammography screening programme; effectiveness, side-effects and cost-effectiveness**

Van Luijt PA, Heijnsdijk EAM, de Koning HJ

Final project report to the Research Council of Norway, 2014

*Summary of methods and results*

This is a model study, where the authors used the MIcro-simulation SCreening ANalysis (MISCAN) model, developed in the 1980s, to model the expected trends in incidence and mortality in Norway following screening implementation under various assumptions. The validity of the different models was assessed through comparison with the observed trends. The same models were used to estimate change in breast cancer mortality, level of overdiagnosis, and cost-effectiveness.

The models simulate a large number of individual life histories that collectively form a population, where a proportion of the population develops preclinical disease (DCIS or early stage invasive breast cancer). Probability distributions for the transition to more advanced disease states and dwelling times[53] at each stage are used to predict the incidence and mortality in the simulated population. From the DCIS state, tumors can also regress. This "disease part" of the model is used to predict the situation with no screening. A "screening" component is then added to the simulated life histories, with modifiable parameters on attendance, screening interval, screening age range, and sensitivity of the screening instruments. For model calibration of the effect of screening on mortality, a 20% reduction was used arguing that this was a conservative estimate from the Swedish RCTs. The models were used to generate two life histories for each simulated individual – one without screening and one with screening. The MISCAN models were first calibrated using life tables and data on incidence in various tumor stages from Dutch screened and unscreened populations. To adapt the models to the NBCSP, life tables and population composition for 2005 was obtained from Statistics Norway. Incidence rates from 1970 to 2009 by age, year and stage, and program attendance by age and year was obtained from the Cancer Registry of Norway. Incidence rates 1970-1990 were used to calibrate the models to a situation without screening. The gradual implementation of NBCSP was modelled by including the proportion of the entire target population who had been invited each year from 1996 to 2004. The observed reduction in breast cancer mortality was calculated as percent change compared to breast cancer mortality in 1990. Observed reduction was then compared to the predicted reduction in the different models. The relative reduction in breast cancer mortality was also translated into numbers needed to screen, calculated as the total number of screening examinations by the number of breast cancer deaths prevented. For all analyses, a population of 10 million women was simulated.

Three models with different assumptions were tested:

Model 1: Simulating NBCSP screening and opportunistic screening (using data from the publication by Lynge et al [117]) as well as a risk ratio of 2.2 for women using hormone therapy, using sales numbers and summary data from the Norwegian Prescription Database to model the extent of use, by age and year. Hormone therapy use was assumed to increase the onset of disease (i.e. the rate of DCIS), but not disease progression (i.e. transition times).

Model 2: In addition to Model 1, another risk factor was added, increasing the number of women developing breast cancer. All women aged 87 and younger in 1997 were modeled to have an additional risk factor for breast cancer in the years 1997-2006 that increased the age-specific hazard with a factor of 1.75.

Model 3: Similar to Model 1, but with an additional assumption of very slow growing tumors, i.e. a large pool of dormant disease. This was achieved by allowing the model to use wider boundaries for the dwell time parameters. For example, dwell time for DCIS was 0.4 years in Model 1 and 4.79 years in Model 3.

[53] *Dwelling time* is the amount of time that a tumor spends in each disease stage before advancing to the next stage.

According to the authors, Model 2 provided the best fit for incidence and mortality combined. Model 1 gave a poor fit both for incidence and mortality; whereas model 3 resulted in a model fit more comparable to that of model 2.For Model 2, the authors also conducted several sensitivity analyses using different levels of screening sensitivity (50 and 100%), attendance (50 and 100%) and screening intervals (1 and 4 years). In these sensitivity analyses, opportunistic screening was not included.
For women aged 55-80 years, the reduction in breast cancer mortality was expected to continue until 2025. In 2025, the predicted reduction in breast cancer mortality was 15% in Model 1, 30% in Model 2 and 25% in Model 3. Model 2 and 3 yielded mortality reductions that are in line what the MISCAN model has estimated in other settings. When including only program screening in Model 2, the reduction in mortality was 20-25%. Changing the attendance rate or the screening interval resulted in estimates from 15% to 30% reduction. The best model fit for increase in incidence over time was observed when the sensitivity in the first screening was assumed to be somewhat lower than the following tests. However, no estimate for mortality reduction was provided for this model. The predicted reduction in breast cancer mortality for women born in 1955 and followed throughout 2055 was 16% in Model 1 and 2, and 13% in Model 3. Assuming an attendance of 80%, the latter predictions were used to estimate number of screenings needed to prevent one death from breast cancer (i.e. efficacy), which were 1676, 1470 and 2612 for Model 1, 2 and 3, respectively.

*Characterization, strengths and limitations*
Van Luijt et al used a modeling approach based on assumptions about the natural history of breast cancer based on several other earlier screening studies including randomized trials. The model utilized population, NBCSP performance data, data on mammography and prescription data from Norwegian sources and compared the model outcomes to observed breast cancer incidence and mortality.

The MISCAN model is developed especially to study screening outcomes and has well-known characteristics as it has been tested in several other settings. The modeling could use empirical Norwegian data for key variables. A model study can test if several different theoretical scenarios are reasonable, i.e. compatible with the observed incidence and mortality trends.

A limitation of all modeling studies is that the results are to a varying degree sensitive to assumptions about phenomena for which there are no empirical data (either because of missing information or to that they in essence are not directly observable). Both Model 2 and 3 include assumptions for which there are no clear-cut theoretical support: Model 2 assumes a risk factor (or a set of risk factors) particular to the Norwegian population which is associated with a substantial risk. Of note is that an alternative explanation is that program performance in NBCSP yields an unusually high incidence. Model 3 assumes longer dwell times than Model 1 and 2, but all models assume a single underlying distribution of lead time.

An estimated use of opportunistic screening was included in all three models, but no distinction could be made between program and non-program screening. These two forms of screening were assumed to have the same effect on breast cancer mortality, and the estimated reductions are thus estimates for the combined effectiveness of program and non-program screening. In the sensitivity analyses including only the number of examinations performed in the NBCSP, the estimated total reduction in mortality was lower due to a lower number of mammography examinations.

The differences between counties in prescreening incidence and hormone therapy use throughout the period, in combination with different timing of screening implementation may have played a role; the models were calibrated using national level data. Furthermore, the model could not incorporate differences in risk factors such as hormone therapy use between attending and non-attending women, which may have led to overestimation of the strength of the additional risk factor in Model 2.

As in other studies of breast cancer mortality following NBCSP implementation, distinction between the effect of screening per se and improved management is not possible in this study, since the observed breast cancer mortality used to validate the models, would be a result of the combined effects of screening and management.

*Conclusions*

The two models that best predicted the increase in incidence over time, gave estimates of 25-30% reduction in breast cancer mortality among women aged 55 to 80 year over the period 1990 to 2025, for program and opportunistic screening combined. For a modelled cohort of women who enter the program in 2005 and are followed throughout life, the number of screenings needed to prevent one death from breast cancer was 1470-2612 in these models. Both models however imply unverifiable assumptions. Model 2 raises the question if program performance in NBCSP has characteristics that drive detection rate upwards compared to other programs. The need to make such assumptions to obtain acceptable model fit may result from the lack of inclusion of county-specific data in the analyses.

### 5.1.2 Studies outside the evaluation

**Breast Cancer Mortality in Participants of the Norwegian Breast Cancer Screening Program**

Hofvind S, Ursin G, Tretli S, Sebuødegård S, Møller B
Cancer 2013;119:3106-12

*Summary of methods and results*

The authors compared incidence-based mortality after the first NBCSP invitation in attending and non-attending women (cohort study). All information, including invitation date, attendance, and breast cancer diagnosis and death, was measured at an individual level. Women were followed from the date of their first screening invitation until the end of 2009 for breast cancer incidence, and to the end of 2010 for breast cancer death. The exposure was ever/never attended and women changed exposure category from unscreened to screened at their first attendance. Women with a breast cancer diagnosis prior to the first invitation were excluded. The results were adjusted for age, period, follow-up and self-selection using Poisson regression. The estimate was multiplied with an external correction factor as an adjustment for self-selection. The correction factor used by the authors was obtained from the Swedish RCTs after comparing breast cancer mortality among non-attendees and non-invited women [118]. Adjustment for county did not influence the results.

After adjustment for age (continuous), period (continuous) and follow-up (3-year categories), the mortality rate ratio was 0.39 (95% CI 0.35-0.44), and with additional correction for self-selection 0.57 (95% CI 0.51-0.64). When multiplied by the compliance rate (84%), the reduction in mortality for invited women was 36%. As a sensitivity analysis, the authors used the lower and upper limits of the confidence interval for the

self-selection correction factor and obtained 0.44 (95% CI, 0.39-0.49) and 0.75 (95% CI, 0.67-0.84), respectively.

*Characterization, strengths and limitations*

The authors applied a classical closed cohort study design in a comparison of attending and non-attending women.

Strengths of the study include individual data, a long follow-up period, and detailed adjustment for the available confounders. Since all women diagnosed with breast cancer in this cohort would benefit from improved management, regardless of attendance, the findings in this study cannot be attributed to such factors.

The major source of systematic error in this study is whether the adjustment for self-selection is sufficient. The method used is based on estimates from the Swedish randomized trials [118], estimating MRR for non-attenders compared to uninvited women, and may or may not be generalizable to the Norwegian Breast Cancer Screening Program. If factors that determine participation differ between the old randomized trials and the NBCSP, and such factors also influence the risk of death from breast cancer, the correction factor for self-selection from the trials would not apply to NBCSP. Potential determinants of participation in NBCSP include the availability of opportunistic screening, separate and more intense screening programs for women with a family history of breast cancer, as well as the dramatic changes in use of hormone therapy. None of these factors were present when the old trials were conducted. Comparing attending and non-attending women will not provide information on the effectiveness of the screening program, a question that would require a comparison of invited and non-invited women.

According to the authors, 38% of attending women reported that they had mammography examination within the last three years before their first invitation from NBCSP, and 64% reported that they had ever used mammography. Since women diagnosed with breast cancer before baseline were excluded from the analysis, only those who were free from breast cancer at previous (opportunistic or clinical) mammography examinations would be included in the study. Assuming that opportunistic screening prior to NBCSP was more frequent among attendees than among non-attendees, this would bias the association away from the null. Conversely, opportunistic screening among the non-attendees after the first invitation would bias the estimates towards the null. The net impact of opportunistic screening would depend on the extent of use in different groups and over time, and also on the effectiveness, both of which are largely unknown.

The one-year difference between the incidence period and follow-up period may have introduced a small lead time bias due to accrual of breast cancer diagnoses in the attending group, leading to underestimation of the contrast in breast cancer mortality between the groups, as discussed by Njor et al [90].

*Conclusion*

Breast cancer mortality rate ratio was 0.57 (95% CI 0.51-0.64) among program attendants compared to non-attendants after correction for self-selection using an external correction factor. Due to the uncertainty of the validity of this correction factor for the NBCSP, the design is not well suited for evaluating the effectiveness of the screening program.

**Effect of Screening Mammography on Breast-Cancer Mortality in Norway**
Kalager M, Zelen M, Langmark F, Adami HO
N Engl J Med 2010;363:1203-10

*Summary of methods and results*
The study design resembles the design used by Olsen et al [105], but Kalager et al included all counties in Norway. The authors compared incidence-based mortality among women living in included and non-included counties during the period 1996-2005, adjusting for the temporal trend of declining breast cancer mortality through comparison with historical control groups from the same counties in the period 1986-1995. Each county contributed a different number of observation years to each of the four groups, depending on the implementation year in each county. Information on breast cancer diagnosis and breast cancer death was measured at an individual level, while person-years were measured at county level. Implementation date in each county was used as a proxy for date of invitation to screening. Breast cancer mortality was first compared between the current and historical groups calculating relative risks. Next, the relative risk for trend in the non-invited counties was subtracted from the relative risk for trend in the invited counties. These comparisons were repeated for women aged 20-49, 50-69 and 70-84 years. Program effectiveness was evaluated in the 50-69 years group, whereas changes in breast cancer mortality among women aged 20-49 and 70-84 years were used as indicators on the potential effect of improved management in counties with screening implementation. Incidence-based mortality was used in all comparisons, meaning that both diagnosis and death would have to occur within the age range and the calendar time when the woman's county was in the given group.

The difference between the relative risks in the invited and non-invited counties over the study period was a reduction of 0.10 (95% CI -0.04 to 0.24) for women 50-69 years. The corresponding difference in rate differences was 2.4 deaths per 100,000 person-years (95% CI, −1.7 to 6.5). An increase of 0.04 in counties with implemented screening was found comparing rate ratios for women 20-49 years (p=1.0), and a reduction of 0.08 for women aged 70-84 years (p=0.09, confidence intervals were not provided for the latter two groups).

*Characterization, strengths and limitations*
The authors followed open cohorts of women in a partially ecologic design combining individual level data on date of breast cancer diagnosis and death with county level data on exposure and person-years. The study design has several similarities with the study by Olsen et al [105], accounting for trends in breast cancer mortality through comparison with regional and historical control groups, but includes all counties at the expense of identical calendar time periods in the exposed group and the regional control group.

The follow-up period in the invited group ranged from 10 years in the pilot counties to two years in the counties where the program was last implemented. Mean follow-up for women with breast cancer was 2.2 years, which is very short given that the median time from diagnosis to death from breast cancer was 5.5 years in the absence of screening [29].

Date of screening invitation was set to the date of program start in each county. Since the roll-out in each county took place over a two-year period, women who got a diagnosis of breast cancer after screening started in their county, but before they received invitation, will be included in the invited group. These women would have clinically detected cancer and thus a higher risk of death from breast cancer than the true post-

invitation breast cancer patients. This misclassification will bias the association towards the null. As for the other studies of breast cancer mortality, opportunistic screening may bias the association in both directions through different mechanisms; in the non-invited groups reducing mortality from cancers that would otherwise be clinically detected within the same study period, and in the invited group reducing mortality through depletion of breast cancer.

Similar to the study by Olsen et al [105], the comparability of the pilot counties, which contribute substantially to the mortality rate in the invited counties, to the remaining counties may be questioned. A valid comparison would require that the trends in breast cancer mortality were similar across the counties in the absence of screening. The non-invited groups were not completely contemporaneous to the invited groups, a situation that also requires that the trends would be linear.

For the first screening round in each county, the person-years for women 50-69 years were divided between the invited group and the non-invited group to reflect the gradual introduction of the program over the two year period. Since the breast cancer diagnoses were not divided accordingly, the breast cancer mortality in the invited group will be overestimated, and the rate in the non-invited group underestimated.

Attention is given to the possibility that the estimated effectiveness of the program may not be separated from that of improved management following the establishment of breast diagnostic centers and multidisciplinary teams. Although a certain effect of improved management is plausible, it should be noted that changes in mortality among women aged 20-49 and 70-84 years may also reflect changes in breast cancer incidence rates, and that systematic differences between the counties could influence the estimates in unpredictable ways also in these age groups. In addition, a significant proportion of women aged 70-84 years in the study group had been invited for program screening before turning 70 years. Due to lead time, cancers that would arise clinically after 70 years and cause death would be excluded from this age group if they were detected at screening before age 70.

*Conclusions*

In this semi-ecologic study, the estimated difference in rate ratios attributed to screening implementation was 0.10 (95% CI -0.04 to 0.24) for women aged 50-69 years. The absolute difference in breast cancer mortality rates was a reduction of 2.4 deaths per 100,000 person-years (95% CI, −1.7 to 6.5). The use of three control groups allowed correction for historical trends, but systematic differences between screening and control counties may distort the comparison in directions that are hard to predict. A widespread non-program mammography use and lack of individual data on invitation date may dilute any screening effect. Duration of both the exposure period and the follow-up period may be too short to provide information on the long-term effectiveness of a fully implemented screening program. Finally, as in several other studies of NBCSP effectiveness, it is hard to distinguish any effect of screening from effects of the concomitant organization of multidisciplinary specialized teams.

**Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database.**

Autier P, Boniol M, Gavin A, Vatten LJ.
BMJ. 2011 Jul 28;343:d4411

*Summary of methods and results*

The authors compared breast cancer mortality trends in three pairs of European countries to disentangle the impact of mammography screening versus modern treatment and improved management on the decline in breast cancer mortality in the countries. The choice of pairs was based on similarities in health care system and treatment, but different durations of organized screening program. One of the pairs was Norway versus Sweden. All data were measured at an aggregated level and were obtained from the WHO mortality database. Total decline in breast cancer mortality over the period 1989-2006 and year of inflexion point were compared between Norway and Sweden. Poisson regression and join-point regression were used to calculate annual percent change, total percent change and to identify the join-point year in each country.

The Swedish mammography screening program was implemented from 1986 and became nation-wide in 1997 (i.e. approximately 10 years before Norway). The decline in breast cancer mortality from 1989 to 2006 was 16% in Sweden and 24.1% in Norway. These numbers are not directly comparable, since mortality was higher in Sweden than in Norway in 1989. In Sweden there was no inflexion point, which indicates a steady decline during the period, whereas in Norway the inflexion point was in 1994. In both countries, the decline was greater for women 40-49 years than for women 50-69 years and 70 years or older.

*Characterization, strengths and limitations*

The study is based on ecologic comparisons of trend in total (not incidence-based) breast cancer mortality. No estimates of the effectiveness of the NBCSP were calculated.

Ecologic studies are in general not suited for causal inference [98], since association at an ecologic level need not reflect the association at an individual level (ecologic bias, ecologic fallacy). Given the gradual implementation of the NBCSP in combination with the ecologic design, there is substantial misclassification of exposure for women who die from breast cancer. Most deaths occurring during the first years after screening was implemented will be due to breast cancer diagnosed before screening was implemented.

Furthermore, change in breast cancer mortality for all ages combined will dilute any reduction resulting from program implementation, since only deaths from breast cancer diagnosed at age 50 years or later may be avoided through a program starting at age 50 years. Another limitation is that the trend in breast cancer mortality over the same period in the absence of screening implementation is unknown. Although breast cancer mortality remained constant over several decades before the mid-nineties, continued constant rates in the absence of screening may not be realistic in a situation with increasing incidence. Both improved treatment and opportunistic screening may have contributed to the decline in breast cancer mortality before NBCSP implementation. Comparison between countries also carries a high risk of ecologic bias, since the number of study units is low and there may be other differences between the countries than the duration of the screening programs.

*Conclusions*
This ecologic trend study indicates that breast cancer mortality was declining before the start of NBCSP. However, it may not be inferred that the NBCSP has not influenced breast cancer mortality among women invited for screening.

**Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database**
Autier P, Boniol M, La Vecchia C, Vatten L, Gavin A, Héry C, Heanue M
BMJ. 2010 Aug 11;341:c3620

**Changes in breast cancer incidence and mortality in middle-aged and elderly women in 28 countries with Caucasian majority populations.**
Ann Oncol 2008;19:1009-18.
Héry C, Ferlay J, Boniol M, Autier P.

**Quantification of changes in breast cancer incidence and mortality since 1990 in 35 countries with Caucasian-majority populations.**
Héry C, Ferlay J, Boniol M, Autier P.
Ann Oncol. 2008 Jun;19(6):1187-94.

*Summary of methods and results*
All three studies assess changes in breast cancer mortality and incidence over time in European countries, and do not to evaluate the effectiveness of mammography screening programs per se. Since the studies overlap considerably in data sources, study period, methods and outcomes, a common summary will be given.

Incidence and mortality rates were obtained for each country and year from WHO databases and other public sources, standardized according to a European or World population. Join-point regression and log-linear regression is used to calculate join-points and annual percent change (APC) in each country.

In Norway, incidence was increasing over the entire period 1960-2005 for women 50-69 years with join-points in 1975, 1993 and 1997; APC=1.9 from 1960 to 1975, followed by APC=0.6 until 1993, APC=11.4 from 1993 to 1997, and APC=1.9 from 1997 to 2005. The increase in incidence for the age group 50-69 was greater than in many other Western countries. For women 70 years or older, there was an increase in incidence (APC=1.7) until 1988, which was the only join-point, and thereafter a decline in incidence (APC= -0.6). Most other countries experienced a continued increase for women in the 70 years or older age group throughout the entire study period. The total incidence increase from 1990 to 2002 was 8.5% in women 35-49 years and 105.3% in women 50-69 years, and the decline for women 70 years or older was -6.8%.

Mortality increased for women 50-69 years with APC=0.9 until 1978 and declined thereafter (APC= -0.7). For women 70+ the increase in mortality lasted until 1993 (APC=0.8), followed by a steeper decline (APC= -1.7). Mortality decreased in all age groups, but more in the youngest group. The total reduction in breast cancer mortality from 199 to 2002 was -28.3%, -13.5% and -13.2%, for women aged 35-49, 50-69, and 70 years or older, respectively. The inflexion point for mortality decline in Norway was later than in many other European countries and the decline in breast cancer mortality in Norway was larger than the median for European countries.

*Characterization, strengths and limitations*

The studies are ecologic trend studies with no defined aim of evaluating the effectiveness of the NBCSP.

The studies provide an overview of development over time in breast cancer occurrence and mortality in different countries and age groups.

The ecologic study design in these studies are not well suited for causal inference about the effect of NBCSP on national trends in breast cancer mortality due to the inability to separate deaths from breast cancer diagnosed before and after invitation for screening, as well as the uncertainty concerning trends in breast cancer mortality in the absence of a screening program.

*Conclusions*

The studies describe increasing breast cancer incidence rates for women in the screening age range and younger and decreasing incidence among older women during the period of NBCSP implementation. Breast cancer mortality declined over the same period in all age groups. Effectiveness of the NBCSP was not assessed.

### 5.1.3 Summary

The included studies present estimates of NBCSP effectiveness, as well as trends in breast cancer mortality over time. A table summarizing the characteristics of each study is provided below. One study also provides data on the extent of non-program mammography use before and in parallel with NBCSP implementation (Lynge et al - not tabulated).

Five studies represent original research with the specific aim to assess whether mammography screening reduces breast cancer mortality in Norway, all of which had incidence-based breast cancer mortality as the outcome.

The studies by Olsen et al [105] and by Kalager et al [119] have very similar designs (the dif-dif design) with partly ecologic data and comparison of incidence-based breast cancer mortality following screening implementation with incidence-based breast cancer mortality in historical and regional control groups [105, 119]. Results are also similar with estimates of 7-11% (RR 0.93, 95% CI 0.77 to 1.12 and RR 0.89, 95% CI 0.71 to 1.12) and 10% (95% CI -4% to 24%) reduction in breast cancer mortality for women invited to screening. Kalager et al also provide an estimate of the absolute risk reduction of 2.4 deaths per 100 000 person-years (95% CI −1.7 to 6.5). The major potential limitations of these studies include misclassification of exposure due to ecologic data and opportunistic screening, differences between the screening and control groups in factors that influence breast cancer mortality, heterogeneity of exposure level in the screening groups, short follow-up and limited statistical precision (especially Olsen et al).

The open cohort study by Weedon-Fekjær et al [120] is based on individual data and is a comparison of breast cancer mortality in invited and not yet invited women, using methods developed specifically to study incidence-based breast cancer mortality and at the same time include as much of the available data as possible. The estimated reduction in breast cancer mortality for women aged 50-79 years is 28% (95% CI 21 to 36%). Numbers needed to invite to prevent one death from breast cancer among women 50-89 years was estimated to 368 (95% CI 266-508). The most important limitation in this study is possible misclassification of exposure due to opportunistic screening

In the simulation study by van Luijt et al [121], individual life histories with and without screening were simulated and compared with ecologic level data to assess model fit. In the two models that provided the best fit for incidence over time, and incidence according to age and relative change in mortality combined, the reduction in breast cancer mortality for women aged 55-80 years due to the combined effects of program and non-program screening was estimated to be 30% and 25% by 2025, respectively. For a simulated cohort of women born in 1955, the estimated life-time reductions in breast cancer mortality with mammography screening were 16% and 13% in the two models, which was translated into an absolute efficacy measure of 1470 and 2612 screening examinations per breast cancer death prevented. The main limitations of these models include the unverifiable and partly uncorroborated assumptions on breast cancer risk factors and lead times necessary to obtain acceptable model fit.

The study by Hofvind et al [122] is a comparison of breast cancer mortality in ever-attending and never-attending women (all invited) using individual level data only. The estimated reduction in breast cancer mortality is 43% (RR 0.57, 95% CI 0.51-0.64) for attending women and 36% for invited women. The study is not primarily a study of the effect of implementing a screening program, but rather a study of the effect of program attendance. Although estimates are adjusted for self-selection, this remains a major source of uncertainty in the study, as well as potential bias due to opportunistic screening.

Four publications are trend studies based on mortality rates from public databases [123-126]. Change in mortality at a national level during different time-periods is examined, but no estimate is reported for the effect of NBCSP on mortality. The ecologic design and the consequently reduced possibility for causal inference is the main limitation of these studies.

One study[117] comparing the extent of non-program mammography use from several sources indicates that there has been extensive use of mammography outside the program. How much of this non-program mammography that can be considered to act as replacement for program screening remains unclear. Thus, the estimated misclassification bias from non-program screening also contains considerable uncertainty.

*Conclusions*

In the studies of incidence-based breast cancer mortality, the studies with the most detailed level of information including use of individual information on invitation date, and the longest follow-up find a larger reduction in breast cancer mortality associated with mammography screening than the studies using county or municipality as proxy for invitation date and with shorter follow-up. In the Dutch simulation study, however, the assumptions that were made to obtain an acceptable model fit to the observed incidence and mortality rate are not recognized in previous literature. The impact of opportunistic screening remains a limitation in all the studies, although the direction and magnitude of the biases may differ between the studies. In some of the studies, the geographic and temporal differences in breast cancer incidence and mortality rates between the comparison groups represent additional limitations of uncertain influence.

The estimated mortality reduction includes 7%, 10%, 28% and 30% in the four studies using the most reliable designs and methods. Two studies [105, 119], with the lowest estimates, have a semi-ecologic design with the largest possibility for misclassification and the shortest follow-up period. Of the two remaining studies, the study by van Luijt et al applies assumptions of unknown risk factors and non-verifiable lead times, while the study of Weedon-Fekjær et al includes most of the available data,

uses individual data on exposure, has a long follow-up and the model is sensitivity tested for different assumptions. This study is rated by the steering group to present the most reliable estimates. We therefore consider that the most reliable population level estimate for reduction in breast cancer mortality attributable to the implementation of the NBCSP is between 20 and 30% for women aged 50-79 years. These estimates pertain to a situation with a screening program as compared to one without a screening program for women aged 50-69 years and followed to the age of 79. They indicate that the Norwegian program performs at average at the level that is expected in the majority of previous thorough reviews of the mammography screening trials.

Summary table of the identified publications and research reports on mortality following NBCSP implementation

| Study | Design | Exposure | Outcome | Population | Period | Age group | Adjustments | Effect measure | Estimate |
|---|---|---|---|---|---|---|---|---|---|
| *In evaluation portfolio* | | | | | | | | | |
| **Olsen** et al [105] | Cohort/ Dif-dif[1] | NBCSP implementation (municipality) | Incidence based breast cancer mortality | Pilot counties vs latest counties | 1990-2008 | ≥50 | Age, period, region | Rate ratio | 0.93 (95% CI 0.77-1.12 ) |
| **Weedon-Fekjær** et al [120] | Open cohort | Screening invitation (individual) | Incidence based breast cancer mortality | Nation-wide | 1986-2009 | 50-79 | Age, period, cohort, county | Rate ratio | 0.72 (95% CI 0.64-0.79) |
| **van Luijt** et al | Simulation study | Screening invitation (individual, simulated) | Incidence based breast cancer mortality | 10 million women simulated | 1970-2009<br>2005-2055 | ≥50 | Not relevant | Reduction in %<br> | 25-30<br>13-16 |
| *Not in evaluation portfolio* | | | | | | | | | |
| **Kalager** et al [119] | Open cohort/ Dif-dif[1] | NBCSP implementation (county) | Incidence based breast cancer mortality | Nation-wide | 1986-2005 | 50-69 | Period, region | Rate ratio difference | 10 (95% CI -4 - 24) |
| **Hofvind** et al [122] | Cohort | Screening attendance (individual) | Incidence based breast cancer mortality | Nation-wide | 1996-2010 | ≥50 | Age, period, self-selection, county | Rate ratio | 0.57 (95% CI 0.51-0.64) |
| **Autier** et al [125] | Ecologic trend | NBCSP implementation (national) | Breast cancer mortality | Nation-wide | 1980-2006 | All ages, 40-49, 50-69, 70-79 | - | TPC[2], APC[3], join-point | - |
| **Autier** et al [124] | Ecologic trend | Calendar year | Breast cancer mortality | Nation-wide | 1980-2006 | All ages, <50, 50-69, ≥70 | - | TPC, APC, join-point | - |
| **Héry** et al [126] | Ecologic trend | Calendar year | Breast cancer mortality | Nation-wide | 1960-2003 | 50-69, ≥70 | - | TPC, APC, join-point | - |
| **Héry** et al [123] | Ecologic trend | Calendar year | Breast cancer mortality | Nation-wide | 1990-2002 | 35-49, 50-69, ≥70 | - | TPC, APC, join-point | - |

[1] Dif-dif design: A design where the effect measure is a double ratio of rates [2] TPC: Total percent change, [3]APC: Annual percent change

## 5.2 Overdiagnosis

### 5.2.1 Studies in the evaluation

Some of the studies sent to the Research Council of Norway as part of the assessment of overdiagnosis in the NBCSP do not contain a direct estimate of overdiagnosis, but rather methodological considerations and background information. These studies will be discussed very briefly in conjunction with studies from the same groups where estimates of overdiagnosis are presented.

**Estimates of overdiagnosis in the Norwegian Breast Cancer Screening Programme**
Duffy SW, Michalopoulos D
Final report to the Research Council of Norway, 2014

*Summary of methods and results*
Overdiagnosis was estimated using two approaches

1) Comparison of the observed incidence during screening ages to the expected incidence in the absence of screening, with subtraction of the estimated post-screening drop in incidence, with correction for lead time for women who had not reached an age where the post-screening drop could be estimated.
2) Comparison of the expected incidence at prevalence screening and at subsequent screenings if no overdiagnosis occurred to the observed incidence at the different screening rounds. The difference between the two situations was considered an estimate of overdiagnosis.

Overdiagnosis was estimated for invasive cancer and DCIS combined, as well as for invasive cancer alone, and expressed as a proportion of the total number of cancers diagnosed among women 50-84 years during the period 1996-2009, i.e. as a proportion of cancers among both invited and not yet invited women, and as a proportion of cancers detected at screening (UK Panel method D).

Breast cancer diagnoses, dates of screening invitations and attendance, and detection mode (not invited, screen detected, interval cancer, non-attending, not invited due to upper age limit, and opted out) were measured at an individual level, whereas population size of the female population in Norway was measured at an ecologic level, by age and calendar year.

Prescreening incidence rates per individual year (1985-1995) and 5-year age groups (50-54 up to 80-84 years) were used to model the expected incidence in the absence of screening. Trends were predicted per age group for three periods; 1996-2000, 2001-2005 and 2006-2009. Log-linear trends were used. Sojourn time (the time from occurrence of mammography detectable breast cancer to detection due to symptoms, i.e. the maximal lead time) and screening sensitivity was estimated using information on interval cancer rates during the first year following each screening round and the expected incidence rates in the absence of screening, as proposed by Day [127]. Estimation of mean sojourn time and sensitivity was done separately in 5-year age groups (50-54, 55-59, 60-64, 65-69 years) and for invasive cancers alone and in combination with DCIS.

By using method 1), overdiagnosis of invasive cancer and DCIS combined was estimated to constitute 5% of all cancers in Norway in women 50-84 years between 1996 and 2009, 8% of cancers in the screening age range (both invited and not yet invited), and 15% of screen-detected cancers. The corresponding results from method 2) were 6%, 8% and 17%, respectively.

For invasive cancer alone, the method 1) gave 2% of cancers in women 50-84 years, 3% of cancers in women 50-69 years and 7% of screen-detected invasive cancers. The second method resulted in an estimate of no overdiagnosis (minus 179 cancers).

*Characterization, strengths and limitations*

This study uses a method specifically designed to obtain full value of the individual data available for this study. Components necessary to produce the estimations of overdiagnosis mentioned above are

- a forecast of breast cancer incidence 1995-2009 from trends 1985-95 based only on age-period specific incidences (no temporal covariates such as hormone treatment, opportunistic screening or county-specific dates of screening implementation) and
- an exponential sojourn time distribution fitted from the number of interval cancers during the first year after a screen.

The main limitations of the study, to be specified below, include uncertainty concerning the expected incidence in the absence of screening, disregard of heterogeneity in the sojourn time distribution, misclassification of exposure due to opportunistic screening, residual confounding by county, and lack of accounting for life expectancy when adjusting for lead time. Finally, the robustness of the results to the many model assumptions and calculations were not justified through sensitivity analyses.

The expected incidence from 1996 and onwards in the absence of screening is crucial to the validity of the overdiagnosis estimates, since this was used both to calculate excess incidence during screening age, the drop in incidence among women no longer invited, and the sensitivity and the mean sojourn time. The expected incidence was predicted at a national level based on projection of trends from 1985. Opportunistic screening prior to NBCSP implementation would increase the incidence during this period and therefore lead to a higher estimate of the expected incidence, and in turn underestimation of the level of overdiagnosis. Baseline differences in incidence rates between the counties, due to differences in use of hormone therapy and other risk factors, and the different inclusion times for the counties in NBCSP, are not reported to have been included in the models. The observed incidence during screening is calculated for all counties combined, and not restricted to the counties where screening had been implemented. This may lead to underestimation of the excess incidence associated with screening, and therefore an underestimation of overdiagnosis. Similarly, the post-screening drop is estimated for all counties combined from 1996. As a result, an observed deficit among women who were never invited to screening is subtracted from the excess during screening ages. This will also lead to underestimation of the level of overdiagnosis. The negative and theoretically highly unlikely estimate of overdiagnosis that was obtained for invasive cancer using method 2) indicates that either the expected incidence and/or the estimates of sojourn time and sensitivity are not realistic. This is further supported by the fact that the estimated mean sojourn time for all age groups combined was longer for

invasive cancers alone than for invasive cancer and DCIS in combination. In a separate publication based on the same data [128], discussed in brief below, the research group describes some challenges encountered when trying to predict the expected incidence in the absence of screening by using pre-screening rates.

In the estimation of sojourn time distribution, overdiagnosed cases are avoided by basing the estimation on interval cancers. However, the assumption of identically exponentially distributed sojourn times disregards that the heterogeneity of tumor types could lead to longer-tailed distributions of sojourn time. Furthermore, there are indications that a proportion of tumors classified as interval cancers by the NBCSP are detected at opportunistic screening [129].

The authors used estimates of sojourn time to estimate the number of breast cancers diagnosed during screening ages 1996-2009 that would be become symptomatic after 2009, without accounting for the life expectancies for women in different birth cohorts. This would lead to underestimation of the level of overdiagnosis. The authors ignore the issue of competing risks: a screen-detected diagnosis is an overdiagnosis if the woman dies before the cancer becomes symptomatic, which necessitates the inclusion of residual life expectancy in the calculations. In the present models, this leads to a situation where the number of breast cancers expected to become symptomatic after 2009 (and therefore subtracted from the estimate of overdiagnosis) increases with increasing sojourn time. The model is therefore not compatible with the general expectation that longer sojourn times would lead to more overdiagnosis [91].

Finally, the estimates of overdiagnosis as a proportion of all cancers among women 50-84 years and 50-69 years should not be compared directly to other estimates in this report, since the denominators in this study are derived from a larger population (invited and not yet invited women combined), compared to the other studies where the denominator is restricted to invited women.

*Conclusions*

The authors used an original approach with a combination of observed post-screening incidence drop and adjustment for lead time estimated from incidence rate predictions and observed rates of interval cancer. The estimates of overdiagnosis obtained were 15-17% of screen-detected DCIS and invasive cancers, and 0-7% of screen-detected invasive cancers. Assumptions in trend predictions, sojourn time distribution and lack of accounting for competing risk may have influenced the results substantially. We consider the estimates of overdiagnosis expressed as proportions of all cancer detected in 1996-2009 (range 0-8%) less informative due to the gradual implementation of the screening program.

**Trends in aggregate cancer incidence rates in relation to screening and possible overdiagnosis: A word of caution.**

Duffy SW, Michalopoulos D, Sebuødegård S, Hofvind S.
J Med Screen. 2014; 1: 24-9

Note: Since no estimates of overdiagnosis in the NBCSP were presented in this publication, it will only be commented briefly.

The authors compared the predicted trends in incidence rates in the absence of screening, the observed trends in the presence of screening, and the number of cancers actually detected at screening.

Trends in incidence for women aged 30-89 years from 1976 to 1995 were estimated using Poisson regression. Three models with different ways of adjusting for age, period and cohort were estimated:

1) A discrete age-cohort model using five-year groups of age and time period
2) A model with discrete age and continuous period trend
3) A separate period trend for each five-year age group

The estimated trends were extrapolated to the screening period and compared to the observed incidence rates for women younger than 50 years, 50-69 years and 70 years or older (5-year categories). In the period 1996-2000, the estimated excess number of cancers (difference between observed and expected number of cancers) for women aged 50-59 was greater than the number of cancers detected at screening in the same age group. The authors conclude that this may indicate that extrapolating trends from the pre-screening period to the screening period to estimate the incidence rate in the absence of screening will not necessarily fit well with the true trend in incidence, at least not for all age groups and periods. Similarly, the authors observed a lower incidence rate than expected from extrapolation in women too old to be invited to screening, supporting that trend extrapolation may not fully capture the true changes.

A major limitation of this study is that the gradual implementation of the NBCSP was not taken into account. The lack of restriction to counties where screening had been implemented could easily contribute to the two situations of poor prediction highlighted by the authors. As an example, in 1996-2000, there were 1629 invasive breast cancers diagnosed in all women 50-54 years in Norway according NORDCAN. The number provided by the authors is 1806 (including DCIS). The excess number of cancers compared to the expected number from trend projections was 635 for all counties combined, whereas the number of screen-detected cancers was 480 must be from the pilot counties only since these were the only counties with program screening during this period. If there was a steeper increase incidence than predicted also in the non-pilot counties, this may have contributed to the excess being higher than the number of screen-detected. Both non-program screening and hormone therapy use might have contributed to a steeper than predicted incidence trend in the non-pilot counties.

**Overdiagnosis of breast cancer in the Norwegian Breast Cancer Screening Program estimated by the Norwegian Women and Cancer cohort study.**

Lund E, Mode N, Waaseth M, Thalabard JC.
BMC Cancer. 2013;13:614.

*Summary of methods and results*

The authors estimated overdiagnosis as a proportion of cancers among women 50-79 years in a situation with screening, comparing women who attended screening and those who did not, with adjustment for breast cancer risk factors. The study is based on a subgroup of women in the NOWAC cohort (53 363 of 172 478 women) who completed a questionnaire between 2002 and 2007 at age 52 years or older and lived in counties with a fully implemented screening program (i.e. the prevalence round was completed). All data are at an individual level. The women included were 52-79 years in 2005 when follow-up started and were followed throughout 2010 for breast cancer incidence. Women with a diagnosis of invasive breast cancer or DCIS prior to answering the questionnaire were excluded.

Cumulative incidence rates of invasive breast cancer and DCIS for women who had attended one or more NBCSP mammography examinations (regardless of any additional non-program screening) were compared to rates for two control groups: women who only had non-program screening and women who had never been screened. Estimates were obtained using Poisson regression and adjusted for age (52–55, 56–59, 60–64, 65–69, 70–79 years), use of hormone therapy (current yes/no), parity (0, 1-2, 3-4, 5+), body mass index (<25, 25+), education (primary, secondary, college) and maternal history of breast cancer(yes/no). Women who attended NBCSP constituted the reference group. Thus, the estimates correspond to UK Panel's definition B of overdiagnosis. The rates of invasive cancer and DCIS in the NOWAC subgroup were also compared to the entire NOWAC cohort and to the national rates for 2006-2010.

Overall, 79.2% had ever attended NBCSP, whereas 12.1% had attended only non-program screening and 8.6% had never been screened. The most striking difference in risk factors between the three groups was for hormone therapy use: 25%, 32% and 12% were current users in the three groups, respectively. Never-screened women also tended to be older than women in the two screened groups and had a lower proportion of maternal breast cancer history. Compared to the NBCSP-screened group, the fully adjusted estimates for invasive cancer and DCIS combined were RR 0.82 (95% CI 0.61-1.11) for never-screened women, RR 1.05 (95% CI 0.84-1.26) for non-program screening, and RR 0.97 (95% CI 0.82-1.15) for the two control groups combined. For invasive cancer alone, the fully adjusted estimates were RR 0.93 (95% CI 0.69-1.25), RR 1.04 (95% CI 0.84-1.29), and RR 1.00 (95% CI 0.84-1.20), respectively. Compared to national rates, both the study cohort and the entire NOWAC cohort had higher incidence rates for women 65-69 years and lower for women 70-74 years.

*Characterization, strengths and limitations*

Lund et al used a classical cohort design with exposure information based on a self-administered questionnaire in a randomly selected population where the response rate was 62%. While the responders thus may be selected on several characteristics, analyses for all major outcomes are based on an internal comparison of exposed to non-exposed among the responders. The study design implies that the prevalence round of screening is excluded.

The analyses are based on individual information on mammography exposure and on breast cancer risk factors, which few other studies have achieved. The follow-up includes a post-screening period. The study uses directly observable incidence rates and does not assume anything about underlying tumor characteristics such as growth rate.

The main limitations of the study are the small sample size, self-reported exposure to screening, and exclusion of the prevalence screening.

The comparison of breast cancer incidence among NBCSP-attenders and never-screened women reflects the largest difference in exposure and may therefore be regarded the most informative estimate of overdiagnosis. However, this comparison also reflects the largest difference in terms of selection for exposure. For mammograms outside the program it is difficult to establish if they were due to symptoms or true screening mammograms. The number of never-screened women in the cohort was low (n=4599), and only 57 cases of breast cancer were diagnosed during follow-up in this group. As a result, the precision of the estimates was reduced.

Self-reported mammography use defined the exposure and there may be some misclassification between screening exposure and mammograms induced by symptoms. A systematic misclassification of mammograms indicated by symptoms as screening mammograms would overestimate overdiagnosis. However, questions were specifically asking for participation in NBCSP and a re-iteration of the questionnaire indicated a high reproducibility of the answers.

The prevalence screening could not be included in the estimation since the participants had to have at least one invitation to NBCSP to answer the questionnaire on mammography screening in a valid manner. Exclusion of the prevalence screening would lead to underestimation of overdiagnosis. Exclusion of breast cancer diagnoses before start of follow-up would also lead to underestimation of overdiagnosis from incident screening rounds, since the attending group would only consist of women who did not get a breast cancer diagnosis at any of the previous screening rounds.

Although adjustment for the major breast cancer risk factors would reduce the influence of self-selection, residual confounding by unmeasured factors cannot be excluded. There may also be residual confounding by factors that were adjusted for. For example, distinction between current and no current use of hormone therapy will not capture previous use, the duration of use, or type of medication, which may also contribute to breast cancer risk [40].

There are some indications that non-program screening may be more frequent in the NOWAC cohort than in Norway in general: Among the NBCSP-non-attenders, the proportion of women who had opportunistic screening only was 58% (12.1/20.8), which is comparable to the 64% of NBCSP-attenders who reported ever use of mammography before their first NBCSP attendance. The comparison of incidence rates between the NOWAC cohort and national figures 2006-2010 show that the NOWAC participants have a higher breast cancer risk during screening ages, despite a slightly lower attendance in the NBCSP (note that the figures for ages 70-74 and 75-79 years are not directly comparable, since a large proportion of women at a national level had never been invited for screening). NOWAC participation may thus to some extent influence the generalizability of the comparison between program attenders and those having a mammogram outside the program.

*Conclusions*
After adjustment for available information on breast cancer risk factors, overdiagnosis among NBCSP attendants was estimated to be 18% (RR 0.82, 95% CI 0.61 to 1.11) of cancers detected among attendants from age 50 to 79 years. Given the modest statistical precision of the study and the exclusion of the prevalence round, the estimate of overdiagnosis is compatible with most other estimates of overdiagnosis in the NBCSP. The adjusted results indicate that information on breast cancer risk factors is important in observational studies of overdiagnosis.

**Overdiagnosis in the Norwegian Breast Cancer Screening Program – estimation based on record linkage and questionnaire information in the Norwegian Women and Cancer study**
Lund E, Nakamura A, Mode N, Kumle M, Thalabard JC
Submitted manuscript

*Summary of methods and results*
The authors estimated overdiagnosis as a proportion of cancers among women 50-79 years in a situation with screening, comparing cumulative incidence rates among women who attended NBCSP and those who did not. The estimate includes both invasive cancer and DCIS. The design is similar to the previously discussed study by the same research group, but is larger, based on 125 102 of the 172 748 NOWAC participants, aged 49-79 during the period 2005-2011 and exposure to screening is based on program data rather than self-reports. All data were measured at an individual level. Women with a cancer diagnosis prior to 2005 were excluded. Attending women were all women who ever had a NBCSP mammogram, whereas non-attending women were those who had mammograms only outside the program or who never had a mammogram. Women could change status during follow-up from non-attending to attending. Ninety-one percent attended the NBCSP at least once. The prevalence screening was restricted to women aged 49-52 years, incidence screening rounds to the age groups 53-55, 56-59, 60-64 and 65-69, whereas incidence rates in the post-screening period was calculated for women aged 70-74 and 75-79 years. Age-adjusted relative risks were calculated using the Mantel-Haenszel method. Analyses were repeated for women with (n= 6873) and without a maternal history of breast cancer.

There were 2185 invasive cancers and 319 DCIS during follow-up. The proportion of DCIS was higher among women attending the NBCSP compared to those not attending (13.1 versus 10.3%, respectively). The cumulative breast cancer incidence rate was higher in the NOWAC cohort compared to the national data for the period 2007-2011. The cumulative incidence rate for attending women was 9.8%, compared to 10.4% among those not attending. Most of the excess risk in those not attending could be attributed to a higher proportion of women with maternal history of breast cancer. For women without maternal history of breast cancer, NBCSP attendants had a 7% higher breast cancer risk than women who never had a mammogram (corresponding to UK Panel's definition A).

*Characterization, strengths and limitations*
As in the previous study from the same group, they used a classical cohort design in women invited from a random population sample and participating in a questionnaire study. Self-

reported data from questionnaires was however in this study combined with program information about participation from the NBCSP.

The study was based on individual information on screening participation from the NBCSP database and could account for individual information of maternal breast cancer. The study allowed for a post-screening follow-up. The prevalence round was restricted to women who entered the screening program when reaching the lower age limit. In combination with using cumulative incidence rates for women in different age intervals, this design estimates overdiagnosis for women a cohort of women who enter the program when reaching the lower age limit and are followed throughout age 79 years.

The main limitations of this study include limited statistical power, absence of adjustment for screening intensity, and limited adjustment for breast cancer risk factors.

The number of women in the cohort who never had a mammogram is not provided, but must be small, since the 9% non-attenders included both women with non-program mammography and women with no mammograms. Confidence intervals for the estimates and for the age-specific rates were not provided.

The NBCSP data provide an account for more details of exposure to screening, but this was not utilized and such analyses would have been complex.

With the exception of the age standardization through cumulative incidence rates and the separate analyses depending on maternal history of breast cancer, there is no control for other risk factors such as hormone therapy use and obesity. If the risk factor distribution is similar to that in the smaller NOWAC study, this would lead to overestimation of the level of overdiagnosis.

As reasoned for the previous study, the restriction of the study to the NOWAC participants may influence the generalizability of a comparison between the attenders to the NBCSP and those having a mammogram outside the program. Compared to the previously discussed study by the same research group, also based on a sample from the NOWAC cohort, the proportion who ever attended NBCSP is higher (91% versus 79.2%). The difference may be due to disagreement between self-reported and register data or to increasing attendance over time. The pronounced drop in incidence rates for women 70-79 years in the non-attending groups indicates a substantial amount of non-program screening in this group, also supported by a higher proportion of DCIS among the non-attenders compared to the national level [122].

*Conclusions*

The estimate of overdiagnosis of 7% for attendants compared to never attendants in this study is smaller than in the study using a smaller NOWAC sample (discussed above). No measure of precision was provided, but the limited number of never-screened women would indicate that the statistical uncertainty may be considerable. Self-selection remains a limitation, despite restriction to women without maternal history of breast cancer.

The estimated level of overdiagnosis is less than in the UK Panel evaluation models and in the publication from the Cancer Registry of Norway. However, the relatively small group of non-screened women implies a limited statistical precision, so the estimate could be considered to be compatible with a broad range of estimates from this evaluation.

**Overdiagnosis of breast cancer after 14 years of mammography screening.**

Zahl PH, Mæhlen J.

Tidsskr Nor Laegeforen. 2012 Feb 21;132(4):414-7.

NOTE: The information on methods and results provided below is partly based on written communication with the authors, since we considered the information provided in the publication insufficient for evaluation.

*Summary of methods and results*

The authors estimated overdiagnosis among women 50-69 years in the four pilot counties during the period 1998-2009. Information on the number of breast cancer and DCIS diagnoses detected at screening and outside the program per year, as well as person-years, were measured at an ecologic level. For invasive cancer, overdiagnosis was estimated as a proportion of breast cancers for women 50-69 years in a situation without screening and as an absolute number. In addition, for invasive cancer and DCIS combined, the absolute number was estimated. Women aged 70 and 71 years, who were invited to a second screening round in 1998 and 1999 [34], were excluded.

Poisson regression was used to estimate annual percent change in incidence rates of invasive cancer for women 40-79 years during the period 1991-2009, with adjustment for age in 5-year categories and prevalence screening and subsequent screening rounds as binary variables. Next, trends during the period 1991-2009 were examined separately for women aged 40-49 and 70-79 years, whereas trends for women 50-69, 50-59, 60-69 and 70-74 years were restricted to the period 1998-2009. For women aged 50-69 years, the incidence rates during the screening period 1998-2009 were compared to the incidence rates 1991-1994 to estimate the excess incidence.

For all age groups combined, the incidence increase over the period was not statistically significant after adjusting for screening implementation in 1996 and subsequent screening rounds. For women aged 40-49 years, the annual percent increase in incidence 1991-2009 was 0.2% (95% CI -0.4 to 0.9). For women aged 70-79, the annual percent decline over the same period was -1.3 (95% CI -1.9 to -0.6). The incidence during the subsequent screening rounds 1998-2009 was decreasing both for women aged 50-69 years (APC -0.7, 95% CI -1.36 to 0.04) and for women aged 70-74 years (APC -2.3, 95% CI -4.2 to -0.4). The increase in incidence from 1991-1994 to 1998-2009 was found to be 50% for women aged 50-69 years, whereas the total decline for women aged 70-74 years was found to be 7% after adjustment for underlying trend. The post-screening decline was considered as negligible by the authors, and the level of overdiagnosis was thus estimated as 50%, under the assumption of no underlying increase in incidence in the age group 50-69 years. For 2009, this was estimated to correspond to 500 cases of invasive breast cancer at a national level. Based on the incidence rate of DCIS in the pilot counties for women aged 50-69 years 1996-2008, the authors predicted that the number of DCIS diagnoses at a national level in 2009 would be 300, all of which were considered as overdiagnoses. The total number of overdiagnoses in 2009 was estimated to be 800, with a 95% CI from 670 to 935.

*Characterization, strengths and limitations*

This is an ecologic study of period changes, not following individual women and their individual risk factors.

The main limitations of the study, to be specified below, include a crude assessment of incidence trends in the absence of screening, use of ecologic data, no subtraction of the post-screening drop in incidence, and direct generalization of the results from the pilot counties to the remaining counties. Statistically insignificant estimates are routinely equated to zero, even when many other values would have been likely. There are no considerations regarding the robustness of the results.

To estimate the causal effect of screening implementation on the excess number of breast cancer diagnoses, optimal adjustment for time period (i.e. underlying trend) is essential, since this will reduce confounding by changing prevalence of breast cancer risk factors like hormone therapy, reproductive factors, overweight/obesity and others. The authors conclude that there would be no statistically significant increase in incidence for women in the screening age range during the period 1991-2009 in the absence of screening. This conclusion seems to be based on a model predicting a linear trend across all age groups with inclusion of binary variables for every two-year period from 1996 and onwards, as a proxy variable for each screening round. The regression coefficients for these binary variables will describe the combined effects of the screening rounds and all other factors influencing the incidence rates.

The dramatic changes in use of hormone therapy described in section 1.5 would be expected to cause a similar change in incidence as breast cancer screening (increase followed by decline) in approximately the same time period, and makes it complicated to separate the contributions from each factor. A manuscript (commented briefly below) from the same research group indicates a substantial increase in risk for women who use combined estrogen and progestin oral preparations. Others have assessed the impact of hormone therapy use on incidence rates in the 50-69 age group to be of similar magnitude as the implementation of screening [39], and many international studies registered dramatic trends in breast cancer incidence concomitant with the rise and fall of hormone therapy use around 2002 [130].

The use of ecologic data leads to misclassification of exposure (invitation to screening), mainly among women 70-79 years. Most of the birth cohorts who were aged 70-79 during the period 1998-2009 were never invited to screening, since they were older than 69 years during the implementation in 1996-1997. These women could therefore not experience any decline in incidence as a consequence of screening. As a result, the decline in incidence for women 70-74 and 70-79 years in this study would underestimate the drop in incidence for women who leave the screening program due to the age limit, which may explain why the authors found this too small to be included in the estimate. In consequence, the level of overdiagnosis would be overestimated.

Misclassification of exposure through opportunistic screening before NBCSP implementation in the pilot counties may have led to a smaller increase in incidence in the 50-69 than what would be observed in a completely unscreened population and thus underestimate the level of overdiagnosis, whereas concurrent opportunistic screening would bias the estimate in the opposite direction. Opportunistic screening in women 70 years or older who have left the program could result in underestimation of the post-screening deficit and thus overestimate the level of overdiagnosis.

The authors used observations from the pilot counties to estimate overdiagnosis for the entire population of Norwegian women. Women in the pilot counties differ from women in the other counties in several aspects, such as a higher level of hormone therapy use, higher breast cancer incidence and more urban areas. The authors present data that support the

existence of such differences, including the incidence rates for DCIS in the pilot counties and the remaining counties during the study period. The generalization from the pilot counties to the entire country may also have led to overestimation of the level of overdiagnosis.

When comparing the estimate to that in other studies of overdiagnosis in the NBCSP, it should be noted that the authors used a smaller denominator than most other studies. In this study, overdiagnosis is estimated as a proportion of breast cancers detected among women 50-69 years in a situation with no screening. In most other studies of overdiagnosis in the NBCSP, however, overdiagnosis is expressed as a proportion of breast cancers detected among women 50-79 years or 50 years and older (life-long follow-up), either in the absence or presence of screening. The different choice of denominator also contributes to the differences between the estimates.

*Conclusions*

In this ecologic trend study, overdiagnosis of invasive cancers was estimated as an excess of 50% among women in the screening age range, compared to a situation without screening, corresponding to an excess of 500 invasive breast cancers per year. When including DCIS the estimated number of overdiagnoses was 800 (95% CI 670 to 935). The validity of this number is limited by the crude methods used to estimate trends in the absence of screening, and failure to account for the post-screening incidence drop.

**Breast cancer incidence and menopausal hormone therapy in Norway from 2004-2009. A register based cohort study.**

Suhrke P, Zahl PH

Manuscript accepted for publication in Cancer Medicine

NOTE: The information on methods and results provided below is partly based on written communication with the authors, since we considered the information provided in the manuscript insufficient for evaluation.

*Summary of design, methods and results*

The authors investigated the risk of breast cancer according to use of hormone therapy in a cohort based on data linkage between the Cancer Registry of Norway, the Norwegian Prescription Database, Statistics Norway, and the Medical Birth Registry of Norway. All information was measured at an individual level. Women who were 50-65 years old in 2006 were included (born 1941-1956). After excluding 84 women with invasive breast cancer and four women with DCIS detected after start of follow-up but before the woman received her first invitation for screening, 449 717 women were included in the analyses.

Systemic hormone therapy was categorized into estrogen and progesterone in combination, estrogen only and tibolone. The total amount of use during 2004 and 2005 was categorized into no prescriptions, 1-180 defined daily doses (DDD), 181-365 DDD, and ≥365 DDD. Women were followed from 01.01.2006 to 31.12.2009. Risk of invasive breast cancer and DCIS was estimated in Cox regression models with adjustment for age, parity, screening attendance (ever/never during the period 2004-2009) and start of hormone therapy during the period 2006-2009 (yes/no).

83% of the study population had attended the screening program at least once. 26.5% had at least one prescription of hormone therapy in 2004-2005, and 14% had at least one

prescription of combined estrogen and progesterone. Women with prescriptions of ≥365 DDD of estrogen and progesterone in combinations constituted 8.4% of the study population and had a hazard ratio (HR) of 2.06 (95% CI 1.90 to 2.24). Prescription of ≥365 DDD of Tibolone was also associated with increased risk of invasive breast cancer (HR 1.23, 95% CI 1.01 to 1.51), whereas estrogen alone was not (HR 1.03, 95% CI 0.85 to 1.25). Combination therapy with prescriptions of ≥365 DDD was associated with a higher risk of invasive lobular cancer (HR 3.10, 95% CI 2.51 to 3.81) than invasive ductal cancer (HR 1.94, 95% CI 1.78 to 2.12) and DCIS (HR 1.61, 95% CI 1.28 to 2.02). The estimated population attributable proportion of breast cancer was 8.2% for combined estrogen and progesterone use of ≥365 DDD over a two-year period, corresponding to 90 breast cancers in 2006.

Women who attended screening had a higher risk of invasive cancer (HR 1.15, 95% CI 1.05 to 1.24) and DCIS (HR 3.32, 95% CI 2.37 to 4.65) compared to women who did not attend. These estimates were not adjusted for other factors.

*Characterization, strengths and limitations*

The major strength of the study is the use of individual data for all variables in the analyses. However, not even individual data on hormone therapy prescription may not fully reflect actual use of hormone therapy. In addition, use prior to 2004 could not be taken into account, and two years of exposure may not capture the risk associated with long-term use.

Apart from the 88 women with breast cancer detected before screening invitation, the authors state in communication with the steering committee that prevalent breast cancer diagnoses were not included. It is not clear if *women* with prevalent breast cancer were included and followed for occurrence of additional (incident) breast cancers or if the women were entirely excluded from the study. According to NORDCAN, the number of prevalent breast cancer cases among women aged 50-64 years in 2005 was 11 349. Women with known breast cancer may be less likely to use hormone therapy, and inclusion of these women may therefore introduce bias.

Women who had all their births before 1967 would not be included in the Medical Birth Registry, and parity would be missing for these women. These women were included as nulliparous in the analyses, and the authors have informed that they constituted six percent of the total study sample. There could also be substantial threats to the internal validity of the study from factors such as age at first birth, obesity, familial risk of breast cancer, and other breast cancer risk factors.

It is not specified whether women were censored at death and emigration. Adjustment for screening attendance and for start of hormone therapy use from 2006 or later would introduce bias, since both screening attendance and hormone therapy prescription would require that the woman was alive and free from breast cancer at that time-point.

*Conclusion*

Women who had prescriptions of combined estrogen and progesterone treatment were found to have a substantially higher risk of breast cancer than women with no hormone therapy prescriptions. Women who attended the NBCSP were found to have a modestly higher risk of breast cancer than women who did not attend. There are a number of potentially important sources of systematic error in this study that should be taken into account before interpreting the results.

**Lead-Time Models Should Not Be Used to Estimate Overdiagnosis in Cancer Screening**
Zahl PH, Jørgensen KJ, Gøtzsche PC
J Gen Intern Med 2014;29(9):1283-6

Note: Since no original estimates of overdiagnosis in the NBCSP were presented in this publication, it will only be commented briefly.

The authors discussed methods for quantification of overdiagnosis in breast cancer and prostate cancer screening. They argued that the excess-incidence approach should be preferred over models with assumptions on the duration of lead time, since the latter are dependent on the model assumptions of lead time and would underestimate overdiagnosis if screening detects tumors that regress or stop growing. They dismissed the argument that the excess incidence method requires a duration of follow-up that is not yet available, referring to Kalager et al 2012, where they claimed that follow-up after leaving the screening program was 10 years and sufficient to allow for a full compensatory drop in incidence. They also argued that the estimates from that study should be seen as estimates of life-time risk and therefore considered them comparable in magnitude to the estimate of 50% during the screening age range presented by Zahl & Mæhlen 2012 (discussed above).

The authors presented a figure with incidence rates for 70 000 Norwegian women aged 50 years who were invited for their first mammography screening in 1996-2001, and then followed for 10 years, and for 43 000 women aged 60 years in 1996-2001 who were invited for their prevalence screening before age 60 years and followed until 1-5 years after leaving the screening program due to the age limit. These rates were compared to the age-specific rates for all women in Norway during the period 1991-1995. No estimates on overdiagnosis were presented based on these rates.

**Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis**
Zahl PH, Jørgensen KJ, Gøtzsche PC
Br J Cancer 2013;109:2014-2019

Note: Since no original estimates of overdiagnosis in the NBCSP were presented in this publication, it will only be commented briefly.

The aim of the study was to estimate lead time for tumors that are not overdiagnosed (termed clinically relevant tumors by the authors) and compare lead time for clinically relevant tumors with lead time including overdiagnosed tumors (termed model-based lead time by the authors). The authors also compared estimates of overdiagnosis with adjustment for clinical and model-based lead time.

Clinical lead time was estimated under the assumption that maximum clinical lead time is 4 years and on assumptions on how fast incidence rates returns to the background incidence after a screening round. The incidence rates in the first, second, third and fourth year following screening were assumed to be 70%, 30%, 10% and 5% lower than the background incidence, referring to Zahl & Mæhlen, 2012 (communication with the authors disclosed that the rates for the first and second year following screening were derived from

Wang et al [131]). Based on these assumptions, clinical lead time was estimated to be 1.06 years.

Model-based lead time was estimated in two different scenarios assuming that lead time for overdiagnosed cases is 10 or 25 years. The clinical lead time was assumed to be 1 year, and the proportion of overdiagnosis (expressed as a proportion of breast cancers detected during the screening age range in the absence of screening) was varied between 10 and 70%. The estimated model-based lead time varied from 1.8 years to 4.7 years for 10 and 70% overdiagnosis when assuming 10 year lead time for overdiagnosed tumors. When lead time for overdiagnosed tumors was set to 25 years, the total model-based lead time was estimated to be 3.2 and 10.9 years with 10 and 70% overdiagnosis, respectively.

The main limitations of the estimation of lead time in this publication include the strong assumptions that the calculations are based on. The separation of tumors into two groups with lead time of maximum 4 years and lead time longer than the woman's remaining lifetime may be an oversimplification of the heterogeneity of breast cancer.

**Research-based evaluation of the Norwegian mammography screening programme; effectiveness, side-effects and cost-effectiveness**

Van Luijt PA, Heijnsdijk EAM, de Koning HJ

Final project report to the Research Council of Norway, 2014

*Summary of methods and results*

The authors used the MIcro-simulation SCreening ANalysis (MISCAN) model, developed in the 1980s, to model the expected trends in incidence and mortality in Norway following screening implementation under various assumptions. The validity of the different models was assessed through comparison with the observed trends. The same models were used to estimate change in breast cancer mortality, level of overdiagnosis, and cost-effectiveness. A more detailed description of the simulation models were given in section 5.1.1.

Three models with different assumptions were tested:

Model 1: Simulating NBCSP screening and opportunistic screening (using data from the publication by Lynge et al [117]) as well as a relative risk of 2.2 for women using hormone therapy, using sales numbers and summary data from the Norwegian Prescription Database to model the extent of use, by age and year. Hormone therapy use was assumed to increase the onset of disease (i.e. the rate of DCIS), but not disease progression (i.e. transition times).

Model 2: In addition to Model 1, another risk factor was added, increasing the number of women developing breast cancer. All women aged 87 and younger in 1997 were modeled to have an additional risk factor for breast cancer in the years 1997-2006 that increased the age-specific hazard with a factor of 1.75.

Model 3: Similar to Model 1, but with an additional assumption of very slow growing tumors, i.e. a large pool of dormant disease. This was achieved by allowing the model to use wider boundaries for the dwell time parameters. For example, dwell time for DCIS was 0.4 years in Model 1 and 4.79 years in Model 3.

According to the authors, Model 2 provided the best fit for incidence and mortality combined. Model 1 gave a poor fit both for incidence and mortality; whereas model 3 resulted in a model fit more comparable to that of model 2. For Model 2, the authors also conducted sensitivity analyses using different levels of screening sensitivity (50 and 100%), attendance (50 and 100%) and screening intervals (1 and 4 years). In these sensitivity analyses, opportunistic screening was not included.

The estimated level of overdiagnosis included both invasive cancers and DCIS, and was calculated as a proportion of cancers among women aged 50-100 years (population perspective) and 50-70 years (individual perspective) in a situation with screening, corresponding to UK Panel's definitions B and C, respectively.

For the period 2000-2009, overdiagnosis was estimated to constitute 2%, 2% and 11% of the cancers among women aged 50-100 years, in Models 1, 2 and 3, respectively. The corresponding estimates for women aged 50-70 years were 3%, 3% and 19%. The predicted levels for the period 2014-2023 were 2%, 2% and 7% of cancers among women aged 50-100 years in Models 1, 2 and 3, respectively. This corresponded to 3%, 3% and 11% of cancers among women aged 50-70 years. Estimates from the sensitivity analyses based on Model 2 without addition of non-program screening were similar to the main results from Model 2. The remaining sensitivity analyses had no material influence, but were somewhat higher with higher attendance or shorter screening interval.

*Characterization, strengths and limitations*

The MISCAN model is developed especially to study screening outcomes and has well-known characteristics as it has been tested in several other settings. The modeling could use empirical Norwegian data for key variables. A model study can test if several different theoretical scenarios are reasonable, i.e. compatible with the observed incidence and mortality trends.

A limitation of all modeling studies is that the results are to a varying degree sensitive to assumptions about phenomena for which there are no empirical data, either because of missing information or because they in essence are not directly observable. The failure of Model 1 to predict the marked increase in incidence from 1994 may in part be due to the association between hormone therapy and screening attendance (both program and non-program). Furthermore, the differences between counties in prescreening incidence and hormone therapy use throughout the period may also play a role, while the models were calibrated using national level data. We are not aware of any single risk factor that could fit the characteristics for the additional risk factor in Model 2, and smaller increases in several risk factors may be a more plausible explanation.

An estimated use of opportunistic screening was included in all three models, but no distinction was made between program and non-program screening. These two forms of screening were assumed to have the same effect on breast cancer incidence, and the estimates of overdiagnosis are thus estimates for program and non-program screening combined, and results were similar when non-program screening was removed from the model.

The estimates for the period 2014-2023 in particular should be interpreted with caution, since this involves extension of the models outside the range of observed data used to calibrate the models. Since the numerator in the estimates for the age group 50-100 years is larger than in other studies of the age group 50-79 years, the proportion of overdiagnosis will be smaller even if the absolute number of overdiagnosed cases should be similar.

*Conclusions*
Model 2 had the best fit with the observed incidence and yielded a modest estimate of overdiagnosis both from the population (2%) and the individual perspective (3%). However, this model assumes the presence of one or more risk factors particular to the Norwegian population for which we have no clear theoretical candidate. Model 3 assumed longer dwelling times for the earliest stage of breast cancer and gave a somewhat poorer fit than Model 2. The estimated level of overdiagnosis was 7% and 11% in the population and individual perspective, respectively.

### 5.2.2 Studies outside the evaluation

**Overdiagnosis among women attending a population-based mammography screening program.**
Falk RS, Hofvind S, Skaane P, Haldorsen T
Int J Cancer 2013; 133(11): 2756-7.

*Summary of methods and results*
The authors estimated the amount of overdiagnosis for a population of women who follow the screening recommendations (i.e. attend) compared to a population with no screening (corresponding to UK Panel definition A). This is a hypothetical population of women who are invited to a prevalence screening at age 50, who have a total of 10 screenings from 50-69, and who are then followed for the rest of their lives. To estimate this, they compared incidence rates for invited women who attended and invited women who did not attend, using data from the start of the screening program throughout 2009. All data were measured at an individual level.

Since the screening program had not existed for a period sufficient to follow women throughout the entire screening program and for the rest of their lives (the longest follow-up was 14 years), the analyses were restricted to women who had the appropriate age for a hypothetically screened cohort in the different screening phases (prevalence round, subsequent rounds, and post-screening period). For analyses of the prevalence round and the post-period, exposure was categorized ever/never attended, and included all women in the relevant age-groups. Analyses of subsequent rounds were based only on those who attended regularly and those who never attended, whereas those who missed a screening round or more were excluded. Incidence rate ratios for attending compared to non-attending women were calculated for two-year age groups and adjusted for county and calendar year using the Mantel-Haenszel method. The incidence rate ratio for women 80 years and older was set to 1, assuming no post-screening drop in incidence for this age group, since there were few observations in this age group.

Next, the incidence rate ratios were applied to a reference population of women aged 50 years in 2010 who were assigned life expectancy values based on the observed mortality in 2010. Reference incidence rates by age were obtained from three sources with expected minimal influence of opportunistic screening and hormone therapy; a model with adjustment for screening implementation and hormone therapy sales numbers [39], observed rates of invasive cancer according to age in 1980-1984, and observed rates for the birth cohorts 1903-1907. These references were used to calculate the expected number of cases with and without

screening, and the difference was attributed to overdiagnosis among women attending screening. Estimates for invited women were calculated by multiplying the estimate for attending women with compliance proportion of the program which was 0.84.

The estimated proportions of overdiagnosis were quite similar for all three reference populations (model, period and cohort approach). For invited women, the estimates for invasive cancer and DCIS combined were 16.5% (95% CI 10.2 to 22.7), 16.3% (95% CI 9.9 to 22.7) and 13.9% (95% CI 7.9 to 20.1) for the model, period and cohort approach, respectively. The estimates were lower for invasive cancer only and higher for attending women.

To assess the degree of self-selection to screening, the authors compared the incidence rate for women not yet invited and invited women who did not attend during the implementation phase of the program in each county. For women 55 years and older, non-attendees had a higher rate of breast cancer than women not yet invited (291/100,000 versus 239/100,000), but for women younger than 55 years there were no clear differences (incidence rate ratio 0.94, 95% CI 0.83 to1.07).

*Characterization, strengths and limitations*

The authors used an original design to estimate overdiagnosis for a hypothetical cohort of women attending screening and followed throughout the rest of their lives. They combine the characteristics of the expected age-incidence curve for screened women, as described by Boer et al [132] and risk estimation for women in the corresponding two-year age intervals. Next, the estimates are applied to reference populations to estimate cumulative excess incidence.

The design allows estimation of the excess incidence during the screening period as well as the post-screening incidence drop in a situation where following a closed cohort of women through the entire screening program and the post-screening period is not (yet) possible. Individual data ensured precise classification of invitation and attendance status. No assumptions on the duration or distribution of lead time were needed. The use of three sets of reference rates for the age-incidence curve in the absence of screening provides information on the robustness of the method. The authors also present data that enables comparison with results from studies where overdiagnosis has been expressed as proportions of other denominators than the one used in this study (lifetime risk of breast cancer from age 50 in the absence of screening).

The main limitation of the study is the potential influence of self-selection when comparing attending and non-attending women. As an example, it is likely that attendance would depend on use of hormone therapy, since women who use hormone therapy are advised to follow the screening program. The supplementary analysis conducted to assess self-selection indicates that non-attending women 55 years and older may have a higher risk of breast cancer than the general population (women not invited), which would lead to underestimation of overdiagnosis. However, since this comparison could only be made during the implementation phase, this finding may not be applicable to the comparison of women attending regularly and women who never attended.

Misclassification of exposure due to opportunistic screening may influence the results in different manners. Assuming that opportunistic screening prior to the first invitation would be more common among attending than non-attending women, the depletion of cases due to opportunistic screening would be greater among attending women and thus lead to

underestimation of overdiagnosis. This would primarily influence the estimates for women 50-51 years, who were the only age groups used to estimate excess risk from the prevalence screening. Opportunistic screening prior to the first invitation among older women could also have contributed to the difference in rates for women who did not attend compared to women who were not yet invited. If non-attending women go to screening outside the program, the estimate of overdiagnosis could be underestimated. If attending women continue to go to screening after leaving the program at 69 years or have opportunistic screening between the NBCSP rounds, this could lead to overestimation.

If the post-screening drop in incidence continues beyond age 79 years, the assumption of no reduced risk in this age group for attending women would lead to overestimation of the level of overdiagnosis.

It should be noted that the use of remaining lifetime number of breast cancers as the denominator would give a lower proportion of overdiagnosis than in many of the other studies of overdiagnosis in the NBCSP, even if the absolute number of overdiagnosed tumors (the numerator) would be similar.

*Conclusions*

For invited women, the estimates of overdiagnosis for invasive cancer and DCIS combined were 16.5% (95% CI 10.2 to 22.7), 16.3% (95% CI 9.9 to 22.7) and 13.9% (95% CI 7.9 to 20.1) compared to a situation without screening and with life-long follow-up. Despite a design that allows for inclusion of a near fully observed post-screening incidence drop due to individual data, and at the same time use of information for women still in screening, self-selection remains a limitation of the study.

**Overdiagnosis of invasive breast cancer due to mammography screening: Results from the Norwegian screening program**

Kalager M, Adami HO, Bretthauer M, Tamimi RM.
Ann Intern Med 2012;156:491-499.

*Summary of methods and results*

The authors estimated overdiagnosis of invasive breast cancer among women invited for screening as a proportion of breast cancers in situation with no screening. The design and the study population was the same as in the study of breast cancer mortality by the same authors [119]. The authors compared incidence of invasive breast cancer among women living in included and non-included counties during the period 1996-2005, adjusting for the temporal trend of increasing breast cancer incidence through comparison with historical control groups from the same counties in the period 1986-1995. Each county contributed a different number of observation years to each of the four groups, depending on the implementation year in each county. Information on breast cancer diagnosis was measured at an individual level, while person-years were measured at county level. Implementation date in each county was used as a proxy for date of invitation to screening.

Breast cancer incidence was first compared between the current and historical groups calculating relative risks. Next, the relative risk for trend in the invited counties was divided by the relative risk for trend in the not-yet-invited counties. Unlike the mortality study, the comparison is conducted under the unstated assumption that the relative increase in incidence

rate from the historical to the current period is the same in both cohorts (invited and non-invited), i.e. only the multiplicative scale is used.

Two different methods were used to account for lead time; observed deficit among women 70-79 years, and adjustment for a lead time of 2 or 5 years. In the analyses including the observed post-screening incidence drop, overdiagnosis was estimated as a proportion of cases 50-79 years in the absence of screening, corresponding to UK Panel's definition A. Since follow-up in many counties was too short to observe any post-screening deficit, a separate analysis with restriction of the screening group to the pilot counties was also conducted. The second approach of adjusting for lead time involved comparing the incidence rates in the screening groups with those for women aged 2 or 5 years older in the historical control groups combined with exclusion of the prevalence round. In these analyses, overdiagnosis was estimated as a proportion of cancers among women 50-69 years in the absence of screening.

When using the observed post-screening incidence drop to account for lead time and including women in all counties contribute in the screening group, overdiagnosis was estimated to 25% (95% CI 19 to 31). When only women from the pilot counties were considered in the screening group, overdiagnosis was estimated to 18% (95% CI 11 to 24). Estimates with adjustment for 2 and 5 year lead time are 20% (95% CI 13 to 28) and 15% (95% CI 8 to 23), respectively.

*Characterization, strengths and limitations*

The authors followed open cohorts of women in a partially ecologic design combining individual level data on date of breast cancer diagnosis and death with county level data on exposure and person-years. The study design may be considered as a more ecologic variant of the double ratio design used by Olsen et al [105], described in section 5.1.1.

The study design allows correction for trends in breast cancer incidence rates through comparison with regional and historical control groups, but includes all counties at the expense of identical calendar time periods in the exposed group and the regional control group.

The main limitations of the study include the lack of individual information on date of invitation for screening, the short follow-up period, and uncertainty in the adjustment for time trend by the use of geographical and historical control groups.

Ecologic data on invitation date for screening will lead to misclassification of exposure in the current screening group, since some women were diagnosed with breast cancer before they were invited to screening, but after screening started in their county. Such misclassification will be present for all invited cohorts during the first two years after screening starts, and for women 50-52 years in the subsequent screening rounds. According to the web appendix of the study, the authors split the person-years of the first screening round in each county to account for the gradual implementation of the program. If a similar separation of the cases detected during this period was not applied, this would lead to overestimation of the rates in the screening group, and underestimation of the rates in the non-screening group. The result would be an overestimation of the level of overdiagnosis. In Figure 2 in the article, the total incidence rate for women 50-79 years in the pilot counties in 1996 is between 600 and 700 per 100 000 person-years. This is much higher than the 345/100,000 person-years reported by others [104]. In comparison, the detection rate at

screening in 1996-97 was 6.7 cases/1000 screenings for invasive cancer and DCIS combined [133].

There may be extensive misclassification of exposure in the 70-79 years group, which consists of a combination of women who were previously invited and have left the program after 69 years of age and women who were never invited. During the implementation phase of the program, women aged 70 and 71 years in the pilot counties were also invited [34]. As a result, the post-screening deficit will be underestimated and the level of overdiagnosis overestimated. For many counties the follow-up is short, and the prevalence screening round will dominate the incidence rates. For these counties, there is little opportunity to study the post-screening deficit, since only few women will have left the program due to age. This may explain why the estimate is much lower when only the pilot counties are included in the exposed group (18% versus 25% overdiagnosis).

Due to population increase over time, the birth cohorts that constitute the 50-69 years group in the study will be larger than for the 70-79 years group. Differences in cohort size will affect the ratio between excess risk during screening and the post-screening incidence drop and lead to overestimation of overdiagnosis.

Misclassification of exposure due to opportunistic screening in the control groups would lead to underestimation of the time trend in the screening groups and overestimation of the time trend in the non-screening groups and therefore lead to underestimation of the level of overdiagnosis compared to a completely unscreened population. Opportunistic screening in invited women (non-attending or between screening rounds for attending women) or previously invited women (70 years or older) would lead to overestimation of the level of overdiagnosis attributable to NBCSP.

Differences in breast cancer risk factors such as age, hormone therapy use and reproductive pattern between the counties could result in different underlying time trends, in which case adjustment for trends in the geographical control groups would not accurately reflect the trends in the screening groups in the absence of screening. There are indications that differences in hormone therapy use at county level may explain much of the changes in incidence rates during the study period [39]. The small time lag between the groups could also influence the trend estimates, especially if the trends were non-linear.

It should be noted that the estimates do not include DCIS and that the estimates from the different approaches of accounting for lead time should not be compared directly, since they have different denominators.

*Conclusions*

Overdiagnosis of invasive breast cancer was estimated to 25% (95% CI 19 to 31), when expressed as a proportion of cancer among women aged 50-79 in a situation without screening. Restricting the screening group to women in the pilot counties, where more of the post-screening follow-up could be observed due to longer follow-up, gave an estimate of 18% (95% CI 11 to 24). The estimates obtained when assuming a lead time of 2 or 5 years were lower, taken into account that these estimates have a smaller denominator. Comparability between the counties in factors influencing breast cancer incidence, misclassification of exposure in the post-screening period, short follow-up and lack of inclusion of DCIS may have influenced the estimates.

**Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends.**

Jørgensen KJ, Gøtzsche PC.

BMJ. 2009 Jul 9;339:b2587.

*Summary of methods and results*

The authors conducted a review of studies reporting incidence rates before and after implementation of a mammography screening program in different countries and estimated overdiagnosis based on the reported rates. Thus, all data were at an aggregated level. The included studies were from Norway (the pilot counties), United Kingdom, parts of Canada, parts of Australia, and Sweden. Only the estimation of overdiagnosis in Norway will be discussed in detail here, but the method and results are similar for the other countries included. Incidence rates for the pilot counties were retrieved from the following studies: Hofvind et al 2006 [104] and Zahl et al 2004 [134].

Information on incidence rates was available for invasive breast cancer only, and not for DCIS. Linear regression was used to estimate prescreening (1980-1994) and post-screening (2000-2006) trends in incidence for women 50-69 years and 70-79 years. The prescreening trend was extrapolated to the last year of the post-screening period to estimate the expected incidence rate in the absence of screening.

Overdiagnosis was estimated as the observed incidence rate in 2006 divided by the expected incidence rate for women aged 50-69 years, after subtraction of the post-screening drop in incidence among women 70-79 years. The post-screening incidence drop among older women was estimated using the same method, and the deficit was subtracted from the estimated rate of overdiagnosis after compensating for the differences in population size at different ages. Under the assumption that DCIS would represent 10% of the tumors diagnosed in a population invited for screening, the estimated level of overdiagnosis was recalculated dividing the observed incidence in the screening period by 0.9.

For the pilot counties, the increase in invasive breast cancer was estimated as 42% above expected rates (observed 303/100 000 person-years versus expected 213/100 000 person-years), or 90 additional breast cancers per 100 000 women per year in the last observation year. Among women aged 70-79, incidence rates were 15% lower in the post-screening period (246/100 000 person-years) than in the prescreening period (289/100 000 person-years), corresponding to -43/100 000 women per year.

Overdiagnosis of invasive breast cancer in the pilot counties was estimated to 37%. When the assumed 10% DCIS was added to the incidence in the post-screening period for women 50-69 years, the estimated level of overdiagnosis was 52% (RR 1.52, 95% CI 1.36 to 1.70).

*Characteristics, strengths and limitations*

The study is an ecologic trend study with prediction of breast cancer incidence trends after screening implementation for women in the screening age range and older based on prescreening trends.

Since the data are at an ecologic level, there will be misclassification of exposure (invitation for screening). The first two screening rounds were excluded and misclassification of non-invited women as invited in the 50-69 year group will therefore be restricted to the youngest cohorts. Bias from such misclassification will result in underestimation, which will

most likely be small since only a few birth cohorts are involved. For women aged 70-79, there will also be misclassification of exposure, since some of the birth year cohorts (about 30%) will never have been invited to screening. This may lead to overestimation due to too small post-screening deficit.

Accounting for the underlying time trend in incidence in the absence of screening is essential to reduce confounding by factors that influence breast cancer risk apart from screening. Given the dramatic change in hormone therapy use during the implementation phase of the screening program, it is unlikely that the incidence rates for women 50-69 years in the absence of screening would continue to increase in a similar pattern as before screening was implemented. Rather, there are indications that at least some of the incidence increase since 1996 may be attributed to hormone therapy [39].

Opportunistic screening may bias the estimates in many ways. Opportunistic screening before 1996 could have increased the incidence rates used to calculate the time trend and overestimated the time trend in absence of screening (underestimation of overdiagnosis). Opportunistic screening in invited women would have the opposite effect. Opportunistic screening in women 70 years and older could lead to underestimation of the post-screening deficit and overestimation of the level of overdiagnosis.

*Conclusions*

In this ecologic trend study, overdiagnosis of invasive breast cancer and DCIS was estimated to an excess of 52% (RR 1.52, 95% CI 1.36 to 1.70) compared to the number of cases among women aged 50-69 years with no screening. The trend predictions may not have accounted for the influence of hormone therapy and non-program screening in sufficient detail, and the ecologic data level result in substantial misclassification of invitation status in the post-screening group.

### 5.2.3 Summary

Eight publications from 2008 or later that present estimates of overdiagnosis of breast cancer in Norway following the implementation of the Norwegian Breast Cancer Screening Program were identified. Five of these were from the evaluation project groups. Characteristics of each study are summarized in the table below. A study of the association between hormone therapy and breast cancer risk from one of the research groups assigned to estimate overdiagnosis was also included.

Before summarizing the results it is necessary to recollect the several possible definitions of the proportion overdiagnosed. In section 1.2.3 we defined overdiagnosis as follows:

‘Overdiagnosis of breast cancer in the context of a mammography screening program is defined as a breast cancer that would not be detected during the woman’s lifetime in the absence of the program.’ This definition is in principle unambiguous, but not individually verifiable, as discussed above.

It is less obvious what this number should be measured against when calculating a proportion of overdiagnosed. The Independent UK Panel on Breast Cancer Screening provided a helpful classification (described in section 1.5.1) of the most common denominators used by researchers, termed A, B, C, and D and summarized in the Table

below. Of these the Panel recommended B for the *population perspective* (the society's view, including health economic aspects) and C for the *individual perspective* (relevant for the woman who needs to weigh benefits and harms when deciding whether to accept an invitation to participate in a screening program). We note that B will always be smaller than C, reflecting that C disregards the benefit of getting fewer diagnoses in the period after end of screening. However, we support this preference and focus on B and C below, whenever possible.

A fifth denominator in the overdiagnosis proportion is being used by Zahl & Mæhlen from this project: rather than using the observed incidence in women 50-69 years old in a situation with screening they use the estimated incidence in women 50-69 years old in a situation with no screening. Since incidence during screening ages is necessarily lower without screening than with screening the Zahl & Mæhlen proportions are necessarily considerably larger than C when calculated from the same information.

The variation in the reported proportion of overdiagnosis across the studies is considerable, which is partly due to variation in the choice of denominators. For comparison, calculations based on other denominators have therefore been included in the summary table whenever possible. As can be seen from the table, the choice of denominator contributes substantially to the variation in estimates, as does whether or not DCIS is included.

Three studies with individual data compared breast cancer risk in attending and non-attending women and included both invasive cancer and DCIS. The two studies by Lund et al using the NOWAC cohort produced quite different estimates of overdiagnosis for women attending NBCSP compared to never-screened women (18% versus 7% of cancers among women aged 50-79 with screening). The difference may be due lack of precision, self-reported versus program-reported attendance, variation in self-selection and adjustment for breast cancer risk factors. An additional limitation in the smallest of these studies is that the prevalence screening could not be included. Sørum Falk et al used nation-wide data and an original approach to estimate life-time risk of overdiagnosis for a hypothetical cohort of attending women. Attendance was estimated to give a life-time risk of overdiagnosis with point estimates ranging from 16.5 to 19.6%, translating into 13.9-16.5% risk for invited women. The possibility of self-selection cannot be excluded in any of these three studies.

Duffy & Michalopoulos used a mixture of observed data and modeling of lead time to estimate overdiagnosis in two slightly different approaches, yielding estimates of 15-17% overdiagnoses among cancers detected at screening, when including both invasive cancer and DCIS. The extrapolation to the full lead time distribution from the experience during the first year after a screen in connection with lack of accounting for competing risks (the woman's possible death from other causes) has most likely led to an underestimation of overdiagnosis.

van Luijt et al applied the MISCAN model to Norwegian aggregated data, and estimated overdiagnosis under different model assumptions and including both invasive breast cancer and DCIS. The estimates of overdiagnosis were presented as a proportion of breast cancers among women aged 50-70 and 50-100 years in a situation with screening, and ranged from 3 to 11% and from 2 to 7%, respectively, for the period 2014-2023. For the period 2000-2009, the estimates ranged from 3 to 19% for women aged 50-70 years, and from 2 to 11% for women aged 50-100 years. The largest estimates were seen for a model including a long mean transition time from DCIS to invasive cancer (4.79 years). The main limitation in this study is the need for additional assumptions to obtain acceptable model fit compared to previous MISCAN applications.

Kalager et al used a partially ecologic variant of the dif-dif design to study overdiagnosis of invasive breast cancer, and estimated 18% overdiagnosis among women 50-79 years compared to a situation without screening in the analysis with the longest follow-up after screening implementation. Although this allowed correction for increasing incidence in the absence of screening, the method could not account for differences over time in the counties, nor for misclassification due to opportunistic screening. The design only allowed for partial inclusion of the post-screening incidence drop, due to short follow-up and misclassification of exposure among older women.

Zahl & Mæhlen and Jørgensen & Gøtzsche estimated overdiagnosis in ecologic trend studies. Overdiagnosis was expressed as a proportion of breast cancer among women aged 50-69 years in the absence of screening (see above). Zahl & Mæhlen reported 50% overdiagnosis of invasive breast cancer, whereas Jørgensen & Gøtzsche reported 52% including DCIS, and 37% for invasive cancer alone. Partial inclusion of a post-screening incidence drop may have contributed to the lower estimates in the study by Jørgensen & Gøtzsche compared to Zahl & Mæhlen. However, both estimates are probably inflated due to misclassification of exposure in the post-screening period, and failure to account for the impact of factors such as hormone therapy use on breast cancer incidence and non-program screening.

Suhrke & Zahl studied risk of breast cancer among hormone therapy users in a nation-wide cohort of women invited for screening, and found a doubled risk for long term users of combination therapy compared to non-users.

The estimates of overdiagnosis are lower in studies with individual information on invitation status and in studies where the post-screening incidence drop is estimated among previously invited women only, as expected. Studies with adjustment for lead-time as an alternative to inclusion of the post-screening incidence drop also reported lower estimates for those analyses.

The influence of non-program screening and changes in use of hormone therapy remain sources of uncertainty in all the studies, and may have contributed to the challenges associated with estimation of the incidence rates in the absence of screening.

*Conclusions*

The variation in design, analytic approach and use of denominators in the studies both in and outside the evaluation project, complicates the justification for a common estimate, and explains most of the variation in the estimates from 7% to 52%. As discussed above the ecologic and semi-ecologic studies are hampered with misclassification of exposure (invited to screening), and some of them also have short follow-up. The studies using individual data are less prone to misclassification of exposure, and combined with longer follow-up, and inclusion of both invasive cancer and DCIS, are expected to give the most reliable estimates.

We consider the most reliable estimates of overdiagnosis of both invasive cancer and DCIS for women aged 50-79 years compared to a situation without screening (method A) to be within the range of 15% and 25%. For women aged 50-79 years in a situation with screening (method B), we consider the corresponding estimates to be within the range of 15% and 20%. The estimates (C, D and Zahl et al) based on the screening ages only (50-69 years) do not include the period of the compensatory drop after the end of the screening period at age 70 in the denominator. We have approximated those of method C from the information in the publications and found them to be in the range of 20-30%.

Summary table of studies estimating overdiagnosis in the NBCSP

| Study | Design | Data level | Population | Exposure | Outcome | Post-screening drop included | Measure of overdiagnosis (%) A | B | C | D | 50-69 unscreened |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *In evaluation portfolio* | | | | | | | | | | | |
| **Lund** et al [135] | Cohort | Individual | Sample from 17 counties | Attending vs never screened | IBC + DCIS<br>IBC | Yes | *22*<br>*8* | **18**<br>**7** | | | |
| **Lund** et al [136] | Cohort | Individual | Sample from all counties | Attending vs never screened | IBC + DCIS | Yes | **7** | *7* | | | |
| **Zahl** et al [137] | Ecologic trend | Ecological | Pilot counties | Before/after implementation | IBC + DCIS[1]<br>IBC | No | | | | | **50** |
| **van Luijt** et al [121] | Simulation | Individual (simulated) | National | Invited vs non-invited 2014-2023 | IBC + DCIS | Yes | | **2-7** | **3-11** | | |
| | | | | Invited vs non-invited 2000-2009 | IBC + DCIS | Yes | | **3-19** | **2-11** | | |
| **Duffy** et al [138] | Ecologic trend + modeling? | Individual? | National | Before/after implementation | IBC + DCIS<br>IBC | Lead time adj. | | | | **15-17**<br>**0** | |
| *Not in evaluation portfolio* | | | | | | | | | | | |
| **Kalager** et al [139][2] | Open cohort/ Dif-dif[3] | Ecological | National<br>Pilot counties vs non-pilot | Before/after implementation | IBC | Partly | **25**<br>**18** | *20*<br>*15* | | | |
| **Jørgensen** et al [140] | Ecologic trend | Ecological | Pilot counties | Before/after implementation | IBC + DCIS<br>IBC | Partly | | | | | **52**<br>**37** |
| **Falk** et al [83][4] | Cohort | Individual | National | Attending vs non-attending | IBC + DCIS<br>IBC | Yes | **19.4**<br>**13.3** | *16.2* | *29* | | *46* |
| | | | | Invited vs non-invited | IBC + DCIS<br>IBC | | **16.3**<br>**11.2** | *13.6* | *24* | | *38* |

[1]Estimates of invasive cancer and DCIS was provided as an absolute number (800) rather than a proportion.[2]Only estimates using the observed post-screening incidence drop are referred here.[3] Dif-dif design: A design where the effect measure is a double ratio of rates. [4] Only the period approach is referred here to enable conversion to other measures (based on table 3 in Falk et al [83]).

*Numbers in italic were not presented by the authors, but calculated by us to facilitate comparison between the studies*

The numerator in all the measures is the excess incidence for women invited (or attending) compared to women not invited (or non-attending) when followed at least to 79 years of age. The denominator in each measure is

A: Incidence in women 50-79 years in a situation with no screening in Lund et al and Kalager et al, and 50 years or older in Falk et al

B: Incidence in women 50-79 years in a situation with screening Lund et al and Kalager et al, and 50 years or older in Falk et al and van Luijt et al

C: Incidence in women 50-69 years in a situation with screening

D: Incidence of screening-detected breast cancer

50-69 unscreened: Incidence in women 50-69 years in a situation with no screening

## 5.3 Interval cancer

### 5.3.1 Studies in the evaluation

**Overdiagnosis in the Norwegian Breast Cancer Screening Program – estimation based on record linkage and questionnaire information in the Norwegian Women and Cancer study**

Lund E, Nakamura A, Mode N, Kumle M, Thalabard JC
Submitted manuscript

*Summary of methods and results*

The authors estimated interval cancer as a proportion of cancers detected among ever screened women in a sample of 125 102 NOWAC participants, aged 49-79 during the period 2005-2011. All data were measured at an individual level. The estimate includes both invasive cancer and DCIS. Women with a cancer diagnosis prior to 2005 were excluded. Ninety-one percent attended the NBCSP at least once. Age-specific rates of interval cancers per 100 000 person-years were calculated. The prevalence screening was restricted to women aged 49-52 years, incidence screening rounds to the age groups 53-55, 56-59, 60-64 and 65-69. Analyses were restricted to women without a maternal history of breast cancer.

Interval cancers constituted 24% of the number of cancers detected among ever screened women. Forty-four percent of interval cancers were >2 cm at the time of diagnosis, compared to 14% of screening-detected cancers. Forty percent of women with interval cancer had regional lymph node metastases at the time of diagnosis, compared to 18% of women with screening-detected cancers. The age-specific incidence rates for interval cancer were 68, 81, 88, 76, 81, and 17 per 100 000 person-years for women aged 49-52, 53-55, 56-59, 60-64, 65-69, and 70-74 years, respectively.

*Characterization, strengths and limitations*

This cohort study provides descriptive information on age-specific rates of interval cancer and tumor characteristics for the period 2005-2011 for women without a maternal history of breast cancer. Individual level linkage between the NOWAC cohort and the NBCSP database ensures accurate information on screening participation. Still, it cannot be excluded that some of the cancers registered as interval cancers in the NBCSP were detected at opportunistic screening.

The main limitations of this study include limited statistical power and the possible selection of women into the NOWAC cohort who have a higher breast cancer risk and/or more use of mammography than the general Norwegian population, as discussed in sections 5.1.1 and 5.2.1. Hormone therapy increases the risk of interval cancer, both through a general increase in breast cancer risk, increased risk of having a mammogram between the screening rounds, and through an increased risk of a false negative result [38, 141]. The rates of interval cancer may therefore differ between users and non-users of hormone therapy. At a national level, the rates of interval cancer may have changed over time in parallel with the change in hormone therapy use, and should not be regarded as a constant measure.

It should also be noted that no definition of interval cancer was provided by the authors, and that the rates per 100 000 person-years provided in this study are not directly comparable to the measures provided in studies where other numerators were used.

*Conclusions*

In the NOWAC sample, interval cancers constituted 24% of cancers among women with at least one attendance in the NBCSP in 2005-2011, and had less favorable disease characteristics at diagnosis. Both sample size and characteristics of the NOWAC participants should be taken into account before generalization of the findings to all women attending the NBCSP.

**Overdiagnosis of breast cancer after 14 years of mammography screening**

Zahl PH, Mæhlen J.

Tidsskr Nor Laegeforen. 2012 Feb 21;132(4):414-7.

Note: The primary aim of the study was to estimate overdiagnosis in the NBCSP, and a detailed discussion of the study is given in section 5.2.1. Here, only the information concerning interval cancer rates is presented.

As part of the descriptive information in the study, the authors provide incidence rates for interval cancer in the four pilot counties during the period 1998-2005. Information on the number of breast cancer and DCIS diagnoses detected at screening and outside the program per year, as well as person-years, were measured at an ecologic level.

The mean rates of invasive interval cancer after the $2^{nd}$, $3^{rd}$, $4^{th}$ and $5^{th}$ screening rounds were 93, 83, 100 and 89 cases per 100 000 invited women per year, respectively.

**Estimates of overdiagnosis in the Norwegian Breast Cancer Screening Programme**

Duffy SW, Michalopoulos D

Final report to the Research Council of Norway, 2014

Note: The primary aim of the study was to estimate overdiagnosis in the NBCSP, and a detailed discussion of the study is given in section 5.2.1. Here, only the information concerning interval cancer rates is presented.

The aim of the study was to estimate the level of overdiagnosis in the NBCSP. Rates of interval cancer during the first year following screening were used to estimate mean sojourn time and sensitivity of the screening program. For invasive cancer and DCIS combined, the total rate of interval cancer during the first year after screening can be calculated from the descriptive information presented in the report: 0.55/1000 screens across the period 1996-2009.

**Could screening participation bias symptom interpretation? An interview study on women's interpretations of and responses to cancer symptoms between mammography screening rounds.**

Solbjør M, Skolbekken JA, Sætnan AR, Hagen AI, Forsmo S.

BMJ Open. 2012 Nov 12;2(6).

*Summary of methods and results*

The authors studied the experiences of women diagnosed with interval cancer in the NBCSP. 40 out of 173 women diagnosed with interval cancer in two hospitals in Central and Northern Norway were invited to an individual interview. Criteria for invitation were short travel distance to the interview location and time of diagnosis as recent as possible.

The range of time since diagnosis was 6-36 months at the time of the interview. Twenty-six women accepted the invitation.

The interviews were semi structured, lasted 45-60 minutes, and focused on symptoms of breast cancer, views on mammography screening, and reactions to having interval cancer. Interviews were audiotaped, transcribed and read by at least two researchers independently. The themes from the interviews were analyzed and discussed using constant comparison, thematic analysis and meaning interpretation. Delayed help-seeking was defined as waiting more than 3 months.

Twenty-four of the 26 women discovered symptoms (lump) of the cancer themselves, and two were incidental findings. It is not stated if the cancers were false-negative or true interval cancer, but it is stated that most of the women did not mention the possibility of a false-negative mammogram. None of the women had knowledge of other breast cancer symptoms than lumps. Fourteen women contacted the health care service within 2 weeks after discovering a lump, eight women between 2 weeks and 3 months, and two women waited for 6 months. Eight of the women who waited less than 3 months defined themselves as delaying help-seeking, even though they did not meet the criteria set by the authors. Those who acted promptly were convinced that cancer was a possibility, whereas those who waited did so either because they were uncertain about the symptom, they were afraid of being perceived as "whimpering", practical reasons, and reasons related to mammography screening. Those who gave the screening program as a reason for delaying help-seeking fell in two categories: delayed because they recently had a negative mammogram or delayed because they had (or thought they had) an upcoming mammography examination.

*Characterization, strengths and limitations*

The authors used qualitative methods consisting of individual interviews with women diagnosed with interval cancer followed by analysis and discussion of the responses.

The study provides insight into factors determining help-seeking behavior when breast cancer symptoms arise in individuals participating in the NBCSP, as well as into the participants' awareness (or lack of awareness) of the pitfalls of screening programs. The study also demonstrates the occurrence of delayed help-seeking as a consequence of screening, although the study design does not allow a quantification of the duration or frequency of the delay. The definition of delayed help-seeking as waiting 3 months or more was set based on the authors' clinical experience and may be considered as a long delay for symptoms of a potentially life threatening disease.

The women were interviewed up to 3 years after diagnosis. This may affect their views and recall of the time before diagnosis was made. Facing negative feelings such as guilt and anger related to the diagnosis, when interviewed by a stranger, may affect their responses. There may also be a selection of patients with a less aggressive disease (being alive three years after diagnosis), and of more resourceful women than the general population (a larger proportion of these women were working than in the general population of women at the same age). Those who defined themselves as "delayers" did this in retrospect knowing that they had cancer. It is possible that they would not have defined their actions in the same way if a benign condition was detected. The decision to seeking help will not only depend on the characteristics of the women, but also on the nature of the symptoms: a breast lump with a certain set of characteristics would initiate a more immediate reaction than a breast lump with a different set of characteristics.

*Conclusions*

Most of the interviewed women with interval cancer sought help shortly after detection of breast cancer symptoms between screening rounds. Those who postponed help-seeking did so for many reasons, one of them being screening attendance. The findings may not be representative for the experiences of all women with interval cancer.

**Mammography screening and trust: the case of interval breast cancer.**

Solbjør M, Skolbekken JA, Sætnan AR, Hagen AI, Forsmo S.
Soc Sci Med. 2012 Nov;75(10):1746-52.

*Summary of methods and results*

The study is based on the same interviews of the same interval cancer patients as above [142]. Trust was not a primary aim when planning the interviews, but emerged as a research question since the interviewed women remained positive towards the screening program despite being diagnosed with interval cancer.

The main finding was that even though most women felt personally let down by the screening program, they still trusted the program in general (at the population level). The sources of such continued trust were that they all knew women with screening-detected breast cancer who in their view had been saved by the program, and they were convinced that there was statistical evidence of the benefits of the program. They suggested more imaging (technology) to reduce the proved fallibility of the program. All women stated that they planned to continue participating, both for recurrence controls and in the general program when recurrence controls ended.

*Characterization, strengths and limitations*

The study is based on the same qualitative methods and interviews as the study by Solbjør et al described above [143].

The study contributes to an improved understanding of how the NBCSP is perceived by women who have experienced the harms of screening.

Similar to the study on delayed help-seeking, the retrospective nature of the study, as well as the possibility of participant selection are the main limitations. The women that had lost their trust in the program completely might be more reluctant to participate in the study.

Many of the women had a high or intermediate level of education, and many were health care workers. It is likely that this affected their will and ability to see themselves as exceptions in an otherwise well-working program, and that a different sample might have revealed less trust. This may reduce the generalizability of the findings to all women with interval cancer.

*Conclusions*

The study highlights and discusses the seemingly paradoxical situation that women who had no personal benefit from the screening program remained positive towards the program, due to the expectation of an overall population level benefit.

### 5.3.2 Studies outside the evaluation

**Comparing interval breast cancer rates in Norway and North Carolina: results and challenges.**

Hofvind S, Yankaskas BC, Bulliard JL, Klabunde CN, Fracheboud J.
J Med Screen. 2009;16(3):131-9.

*Summary of methods and results*

The authors compared the rates of interval cancers in two screening programs with different organization. Only the results from the NBCSP will be discussed below. The Norwegian data included the four pilot counties from 1996 to 2002. Only interval cancer after the subsequent screening rounds was considered. 151 678 Norwegian women with at least one subsequent screening (mean 1.8 screens) were included. Interval cancer was defined as a diagnosis of invasive breast cancer or DCIS after a negative screen but before the next scheduled screen, or within two years after a negative screen for women who left the program due to the age limit. A negative screen included additional diagnostic work-up with a benign result when this was recommended based on the screening mammogram. There was no censoring at death or emigration.

The rate of screen-detected breast cancer (DCIS and invasive cancer combined) was 5.14 per 1000 screens, and the total rate of interval cancers was 1.81 per 1000 screens. The rate of interval cancers was 0.54 per 1000 screens the first year after screening, and 1.27 per 1000 screens the second year. The proportion of DCIS among all interval cancers was 4.8% during the first year and 7.0% during the second year after screening (rate 0.03 and 0.09 per 1000 screens, respectively). Tumor size and lymph node involvement was similar for interval cancer cases diagnosed during the first and second year after screening.

*Characterization, strengths and limitations*

The authors used an open cohort design including women who attended subsequent screening rounds in the pilot counties.

The main strengths of the study are the individual level of information and the ability to compare rates and characteristics of interval cancers for the first and second year following screening.

Cancer detected at opportunistic screening between the program screening rounds would be classified as interval cancer in the NBCSP database. This may lead to an overestimated rate of interval cancers. The rates of interval cancer may vary over time and between counties. Women who use hormone therapy have a higher risk of interval cancer, both due to a higher risk of false-negative mammography and due to a higher risk of breast cancer in general. Thus it is likely that the peak of hormone therapy use during the study period in this study may have influenced the results.

*Conclusions*

Interval cancers constituted 26% ((1.8 / 1.8 + 5.14) x 100%) of the total number of cancers detected among screening attendants during the two-year period of a subsequent screening round in the pilot counties.

**A pooled analysis of interval cancer rates in six European countries.**

Törnberg S, Kemetli L, Ascunce N, Hofvind S, Anttila A, Seradour B, Paci E, Guldenfels C, Azavedo E, Frigero A, Rodriges V, Ponti A.

Eur J Cancer Prev 2010;19:87–93.

*Summary of methods and results*

The aim of the study was to obtain a pooled estimate for detection rate and interval cancer rate across screening programs in the European Breast Cancer Screening Network and to discuss causes for similarities and differences. Data from Spain (Pamplona), France (Marseille and Strasbourg), Italy (Torino and Florence), Sweden (Stockholm), Finland (Pirkanmaa) and Norway (the pilot counties) were included. From the pilot counties, only the prevalence round and the following screening interval was included (interval cancer after subsequent screening rounds is addressed in Hofvind et al [144]). From the other countries, data from both the prevalence and the subsequent rounds were included, except Finland (only prevalence round). Data were obtained from the screening centers in each program. Interval cancer was defined as a cancer diagnosis after a negative screening test (mammography with or without recall) within a time period equal to the screening interval. The denominator was the number of screening examinations. The rates of interval cancers were compared against detection rate and background incidence, the latter was provided by each center and was the incidence during the period 3-5 years before start of screening. Both invasive cancers and DCIS were included.

The pooled rate of interval cancer was 0.59/1000 screens during the first year after screening (range 0.21-0.73/1000 screens), and 1.26/1000 screens during the second year (range 0.63-1.50/1000 screens). For the pilot counties, the rates were 0.45/1000 and 1.5/1000 screens, respectively. Pooled detection rate in the prevalence round was 6.04/1000 screened (range 4.16-9.10/1000 screened), and for the pilot counties 6.72/1000 screens. The ratio of interval cancer rate and background incidence rate was 0.49 for Norway (background incidence 200 per 100 000) and 0.46 for all countries combined (range 0.26-0.67, background incidence assumed to be 200 per 100 000 person-years).

*Characterization, strengths and limitations*

For Norway's part, with inclusion of the prevalence round in the pilot counties only, the study design consists of a closed cohort followed for a period of 2 years.

As for the study by Hofvind et al discussed above, individual data and distinction between the first and second year following screening represent strengths of the study. The background incidence is estimated from the years before screening implementation, and may not reflect the true incidence in the absence of screening. A comparison of interval cancer rates with incidence rates among not yet invited women will also introduce the possibility of self-selection, since women with interval cancer are screening attendants, and may have a different breast cancer risk than the general population in the absence of screening.

*Conclusions*

Interval cancer constituted 22% of the total number of breast cancers diagnosed among screening attendants during the 2-year period of the first screening round in the pilot counties.

**Prognosis in women with interval breast cancer: population based observational cohort study.**

Kalager M, Tamimi RM, Bretthauer M, Adami HO.
BMJ. 2012 Nov 16;345:e7536.

*Summary of methods and results*

The authors studied prognosis in breast cancer patients diagnosed 1996-2005 at ages 50-72 years in all counties in Norway. Women with interval cancer (n=1816) were compared to women diagnosed before they received their first screening invitation (n=5300). Follow-up for death from breast cancer was until the end of 2006. Interval cancer was defined as a first diagnosis of invasive breast cancer within two years and two months after the last normal mammogram but before the next invitation (if there was a next invitation). Prognosis was compared using life tables (Kaplan-Meier) and Cox regression. The regression models included age (four categories), period, county and time since last screening.

There were no clear differences in prognosis for women with interval cancers and pre-screening cancers (HR 0.98, 95% CI 0.84 to 1.15), and no clear trends according to time since last screening, age at diagnosis or period. Women with interval cancer more often had tumors with lobular histology, larger tumors and less often lymph node involvement compared to women not yet invited. A larger proportion of women with interval cancer had sentinel node biopsy examination.

The rates of interval cancer in this study were 163.4 in the first screening round, 162.5 in the second, 193.3 in the third, and 166.5 in the fourth screening round (all per 100 000 person-years).

*Characterization, strengths and limitations*

The authors used a classical cohort design with individual information on all variables.

Opportunistic screening in both groups would lead to misclassification of exposure. Women registered with interval cancer may have screen-detected cancer if they had opportunistic screening mammography at private centers between two public screening rounds. Women in counties where screening had not yet been implemented may also have screening-detected cancer through opportunistic screening. The magnitude of such misclassification is not known, and the impact of the resulting bias is difficult to assess. Women with interval cancers are screening attendants, and are not completely comparable to the general population of non-invited women with breast cancer. The results are adjusted for age, period and county, and the prognosis is similar both in the unadjusted and adjusted analyses, suggesting that the covariates had little combined influence on the association. There may be differences in use of hormone therapy and other risk factors that are not captured by the included covariates. Hormone therapy users have a higher risk of interval cancer than non-users, and have been encouraged to attend screening according to clinical guidelines.

It should be noted that the rates of interval cancer provided in this publication are much higher than those reported in other publications. It is not stated if the rates are for invited or attending women.

*Conclusions*

There were no differences in prognosis for women with interval cancers and women not yet invited for screening. It is not clear how the combined influence of lead time bias due

to opportunistic screening among not yet invited women and self-selection among women with interval cancer would affect the results.

**Mammographic features and histopathological findings of interval breast cancers.**
Hofvind S, Geller B, Skaane P.
Acta Radiol. 2008 Nov;49(9):975-81.

*Summary of methods and results*
The study aim was to describe radiological and histopathological features of cancers missed at screening and true interval cancers. The study population consisted of 231 women 50-69 years who participated in the prevalence screening between November 1995 and March 1998 and were diagnosed with interval cancer following this screening. Interval cancer was defined as cancer diagnosed within two years after a normal mammogram or an abnormal mammogram with benign results at further assessment. A blinded review of the interval cancers was performed by six radiologists in a consensus meeting. The results of this review were published in Hofvind et al [145]. The screening and diagnostic mammogram, and pathology and surgery reports were used in the review. 80 cancers were defined as missed by all the radiologists, 53 as minimal signs, 82 as true, 16 as occult and 16 cases had no available diagnostic mammogram. Tumor characteristics were compared for the following three groups: missed (n=80, 35%), minimal signs (n=53, 23%) and true/occult (n=98, 42%). Proportions with specific tumor characteristics were compared using chi square tests.

There were no clear differences in radiological characteristics between the missed and minimal sign tumors. A poorly defined mass or asymmetric density was the most common radiological feature at the screening mammogram for both groups. Missed tumors were more often lobular cancers and larger at diagnosis than minimal sign and true interval cancers. A lower proportion of true interval cancers had lymph node involvement at diagnosis compared to missed and minimal sign cancers. There were no clear differences in tumor grade or receptor status between the groups.

*Characterization, strengths and limitations*
The comparison of tumor characteristics was conducted in a cross-sectional design, after retrospective blinded review of clinical and screening mammograms.

The distinction between true and missed (false negative) interval cancers in this study provides a valuable supplement to studies of interval cancers in larger samples where such a distinction is not available. No adjustment for other factors was made in the comparisons. Patient characteristics such as age at diagnosis and use of hormone therapy may influence both the risk of being missed at screening and also histopathological features of the tumor. Some of the cancers classified as true interval cancers could be tumors detected at opportunistic screening, which could affect the proportions of missed and true cancers. The statistical power was limited, and there may be differences between the groups that would have been clearer if a larger sample was examined.

*Conclusions*
Among women diagnosed with interval cancer following the prevalence screening in 1995-1998, 35% of interval cancers were cancers missed at screening, and these cancers tended to be larger when diagnosis was made and more often had a lobular histology.

**Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography-- results from a retrospective review.**
Hoff SR, Abrahamsen AL, Samset JH, Vigeland E, Klepp O, Hofvind S.
Radiology. 2012 Aug;264(2):378-86.

**Missed and true interval and screen-detected breast cancers in a population based screening program.**
Hoff SR, Samset JH, Abrahamsen AL, Vigeland E, Klepp O, Hofvind S.
Acad Radiol. 2011 Apr;18(4):454-60.

The two studies are summarized and discussed together since the most recent of the studies also contain results from the first study.

*Summary of methods and results*
The aim was to compare the proportions and mammographic features of missed cancers at digital and film screening. The authors conducted a review of mammograms for women diagnosed with interval cancers and screening-detected cancers in Vestfold and Møre og Romsdal counties. Vestfold has had digital mammography since the start in 2004, whereas Møre og Romsdal had screen film mammography from the start in 2002 to 2008. The comparison between digital and film screening was therefore also a comparison of the two counties.

The review included 49 interval cancers after the prevalence round with digital screening in Vestfold (2004-2005), and 86 screening-detected cancers at the subsequent screening in 2006-2007. From Møre og Romsdal, 81 interval cancers from both prevalence and the first subsequent screening round and 123 screening-detected cancers from the first and second subsequent round were included. Both DCIS and invasive cancers were included in both counties. The review was performed by four radiologists with more than 5 years of experience. The cancers were classified as missed, minimal signs (further subdivided into actionable and not actionable), and true. Both prior and diagnostic mammograms and pathology reports were used in the review. The missed and minimal signs actionable groups were combined in the analyses. The characteristics of the different breast cancer groups were compared using chi square tests, Fisher exact tests, and t-tests for independent samples.

The counties differed in several screening parameters: Participation, recall, detection of DCIS at screening was higher in Vestfold at both the prevalence and subsequent screenings. The background incidence (1997-2001) was also higher in Vestfold (258 versus 216 per 100 000). Thirty-three percent of the interval cancers after digital mammography, and 30% of the interval cancers after screen film mammography were classified as missed (missed and actionable minimal signs). The corresponding numbers for screening-detected cancers were 20% for digital and 21% for screen film. Among screening-detected cancers there was a higher proportion of DCIS than among interval cancers, but no clear differences according to screen film or digital mammography.

Mammographic features of missed cancer (interval and screening-detected combined) were quite similar to the non-missed screening detected cancers. A mass was by far the most common feature, but also calcification and/or asymmetry were common features. With the exception of size, the masses had similar characteristics (shape, margins and density) for digital and screen film of both missed and non-missed cancers. Size, however, was larger for non-missed screening-detected tumors than for missed

cancers and in both these groups, the identified masses were larger at screen film than at digital mammography.

*Characterization, strengths and limitations*

Cancers missed at screening were investigated in a retrospective review of digital and screen-film mammograms.

The study provides information on the extent and characteristics of missed cancers which will be valuable for quality assurance and improvement of the screening program.

Since the digital screening population and the screen film population of cases came from two separate screening centers in counties with different pre-screening incidence rates, it is not possible to separate the effect of digital versus screen film mammography from the effect of center and county. It is not stated if the radiologists were blinded to whether an individual case was a screen-detected cancer or an interval cancer. No blinding may allow expectations to influence interpretation of the mammograms. For example, knowing that a cancer was detected before the next screening round may make the radiologist evaluate the previous mammogram even more carefully.

Møre og Romsdal had one more screening round than Vestfold, and the interval cancers in this county was a combination of interval cancers after the prevalence and the subsequent screening round, as opposed to Vestfold where the interval cancers were detected after the prevalence round only. Another study from Møre og Romsdal by the same authors indicates that 8% of cancers registered as interval cancers in the NBCSP were detected at opportunistic screening, and that a total of 19.6% of the interval cancers were asymptomatic [141]. If the situation was different in Vestfold (more or less opportunistic screening), this may influence the comparison of the missed/non-missed ratio between digital and screen film mammography and the comparison of tumor characteristics between the groups.

The number of reviewed cancers was limited, and there may be differences between the groups that would be more apparent in a larger data set.

*Conclusions*

The proportions of missed cancers at screening were similar in two counties with different mammography technologies. Combined, 30-33% of interval cancers had been missed at screening, and 20-21% of cancers detected at screening had been missed at the previous screening.

**Screening-detected breast cancers: discordant independent double reading in a population-based screening program.**

Hofvind S, Geller BM, Rosenberg RD, Skaane P.
Radiology. 2009 Dec;253(3):652-60.

*Summary of methods and results*

The aim of the study was to compare cancers from discordant and concordant readings in the NBCSP. 1 059 309 prevalent and subsequent screens for women who participated between 1996 and 2005 were eligible. 5978 DCIS and invasive tumors were detected at these screenings. Mammograms interpreted by radiologists who had read <500 mammograms or worked less than 1 year in NBCSP were excluded, as were mammograms with no double reading, missing information on scoring, laterality, mammograms in women with symptoms and technically inferior mammograms (25 439 screens and 367 cancers). The scoring by each reader was according to the following

classification: 1, normal; 2, probably benign; 3, indeterminate; 4, probably malignant; and 5, malignant. Consensus was required when one reader gave a score of one and the other 2 or more and the reading was considered discordant. If both gave 2 or more, the reading was considered concordant positive, but consensus to decide whether to recall the woman or not was still required. If both gave a score of 1, the reading was considered concordant negative.

The detection rate was 5.4 per 1000 screens, the rate of interval cancers was 1.7 per 1000 screens, and the recall rate was 3.5%. Ninety-seven percent of the mammograms included were screen-film. The median number of readings for each radiologist was 2995 screens (range 275 – 13 395), whereas the recommended number was 5000 and 35% of the 107 included radiologists met this criteria. Overall, 92.6% of the screens were concordant negative, 5.3% were discordant, and 2.1% were concordant positive. Among women who were recalled, equal proportions had discordant and concordant positive readings. Among all screening-detected cancers, 23.6% had discordant readings. The proportion of discordant cancers was not related to the reading volume. The rate of interval cancer was 1.7 per 1000 concordant negative screens (n=1674), 2.9 per 1000 dismissed (not recalled) discordant screens (n=105) and 3.1 per 1000 dismissed concordant positive screens (n=12).

*Characterization, strengths and limitations*

It was not possible to link the interval cancers directly to the lesions that were given a score of 2 or more. It is therefore not possible to quantify the number of false-negatives directly from this study. Some cancers classified as interval cancers in this study may have been detected at opportunistic screening. Use of hormone therapy is associated with breast density, which influences the sensitivity and specificity of mammography [108, 109]. In addition, most mammograms in the study were screen-film mammograms. Both concordance rates per se and the interval cancer risk according to concordance status may have changed with decreased use of hormone therapy and increased use of digital mammography.

*Conclusions*

The rates of interval cancer were higher for women who had discordant or positive mammograms before consensus, but who were not recalled.

### 5.3.3 Summary

The included studies report the frequency or incidence rates of interval cancers, their characteristics and prognosis, as well as the experiences of women with interval cancer.

There are some discrepancies in the published rates of interval cancer across studies. The rate of interval cancer (invasive and DCIS combined) following the prevalence screening round in the pilot counties was 1.95/1000 negative screens and the detection rate at screening was 6.72/1000 screens [146]. Following the subsequent three screening rounds, the rate of interval cancer (invasive and DCIS combined) in the pilot counties was 1.80 per 1000 screens and detection rate at screening 5.14 per 1000 screens [144]. Approximately 70-75% of the interval cancers were diagnosed in the second year following screening. In a study based on nation-wide data until 2005, the rate of interval cancer was 1.7/1000 screens [151]. Most studies are compatible with these rates of interval cancer [136-138]. Another study report that the rates of interval cancer (invasive only) for the entire country varied between 162.5 and 193.3 per 100 000 person-years [147], which is approximately twice as high as in the other publications. This may result from the lack of taking into account that most women contribute two person-years during one screening interval. Information on the occurrence of interval cancer from the included studies is summarized in the table below.

Interval cancers were larger at the time of diagnosis and more frequently had lobular histology, but less frequently lymph node involvement, than cancers diagnosed before screening invitation [147]. Interval cancers that were missed at screening (i.e. false negative) were more frequently lobular tumors and were larger than true interval cancers [148]. Tumor characteristics were similar for cancers detected in the first and second year following screening [144]. The prognosis for women with interval cancer was similar to the prognosis for women with breast cancer detected before screening implementation [147]. Opportunistic screening and self-selection may have influenced the prognosis for these two groups of women.

The proportions of true interval cancers and cancers missed at screening have been investigated in three studies. These studies indicate that 30-35% of the cancers diagnosed in the two-year interval following a screening, could be detected when reviewing the screening mammogram (i.e. were false negative) [148-150]. In comparison, approximately 20% of the screening-detected cancers could also be detected when the previous screening mammogram was reviewed and were false-negative. The rates of interval cancers were higher for women who were not recalled after consensus, but who had a mammogram that was considered abnormal by at least one radiologist [151].

Two qualitative studies of the experiences of women diagnosed with interval cancer indicate that some women postponed help-seeking when they developed breast cancer symptoms, due to the negative mammography examination [143] and that despite being diagnosed with interval cancer, the interviewed women remained positive towards the screening program [142].

Summary table of occurrence of interval cancer in the NBCSP

| Study | Design | Data level | Population | Study period | Cancer | Screening round | Age group | Rate two-year interval | Rate first year | Rate second year | % of cancers in attendants |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Studies in evaluation portfolio* | | | | | | | | | | | |
| **Lund et al [136]** | Cohort | Individual | Sample from all counties | 2005-2010 | IBC + DCIS | Prevalence<br>Subsequent | 49-52<br>53-55<br>56-59<br>60-64<br>65-69<br>70-74 | 68 / 100 000 py<br>81 / 100 000 py<br>88 / 100 000 py<br>76 / 100 000 py<br>81 / 100 000 py<br>17 / 100 000 py | | | 24% |
| **Zahl et al [137]** | Open cohort | Ecological | Pilot counties | 1998-2005 | IBC | Subsequent 2nd<br>Subsequent 3rd<br>Subsequent 4th<br>Subsequent 5th | 50-69<br>50-69<br>50-69<br>50-69 | 93 / 100 000 py<br>83 / 100 000 py<br>100 / 100 000 py<br>89 / 100 000 py | | | |
| **Duffy et al [138]** | Open cohort | Individual | National | 1996-2009 | IBC + DCIS<br>IBC | Prevalence + subsequent | 50-69 | | 0.55 / 1000 screens<br>0.52 / 1000 screens | | |
| *Studies not in evaluation portfolio* | | | | | | | | | | | |
| **Hofvind et al [144]** | Open cohort | Individual | Pilot counties | 1996-2002 | IBC + DCIS | Subsequent, 2nd and 3rd | 50-69 | 1.81 / 1000 screens | 0.54 / 1000 screens | 1.27 / 1000 screens | 26% |
| **Törnberg et al [146]** | Cohort | Individual | Pilot counties | 1996-1997 | IBC + DCIS | Prevalence | 50-69 | 1.95 / 1000 screens | 0.45 / 1000 screens | 1.5 / 1000 screens | 22% |
| **Kalager et al [147]** | Open cohort | Individual | National | 1996-2005 | IBC | Prevalence<br>Subsequent 2nd<br>Subsequent 3rd<br>Subsequent 4th | 50-69<br>50-69<br>50-69<br>50-69 | 163.4 / 100 000 py<br>162.5 / 100 000 py<br>193.3 / 100 000 py<br>166.5 / 100 000 py | | | |
| **Hofvind et al [148]** | Open cohort | Individual | National | 1996-2005 | IBC + DCIS | Prevalence + subsequent | 50-69 | 1.7 / 1000 concordant neg. screens<br>2.9 / 1000 discordant pos. screens<br>3.1 / 1000 concordant pos. screens<br>1.7 / 1000 all screens combined | | | |

[1]py: person-years

## 5.4 Experiences with screening participation and recall examinations

### 5.4.1 Studies in the evaluation

**Informasjon, valg og posisjoner ved mammografiscreening.**
Solbjør M
Book chapter. In: Tjora, A (red) Helsesosiologi. Analyser av helse, sykdom og behandling. Gyldendal Akademisk, Oslo, 2012. In Norwegian.

*Summary of methods and results*
The authors studied of the experiences of women who had participated in the Norwegian Breast Cancer Screening Program. Women who participated for the first time and women who had participated several times were invited for focus group interviews. Women in the latter category were recruited based on previous participation (2003) in a similar study by the same authors.

Thirty-five previously interviewed women who were residents in Nord- and Sør-Trøndelag were contacted, 31 agreed to receive information and invitation for the follow-up study. Twenty-four women accepted (four focus groups). The women that participated in screening for the first time were recruited based on random sampling from the screening program administration system. Forty women from Trondheim and 40 from Bodø were invited, and 14 accepted (three focus groups). The interviews were organized as open conversations according to a predefined interview guide. The groups were asked one question at a time, and a new question was given when conversation on the previous question stopped. Two questions were related to the women's need for information. The interviews were audiotaped and transcribed. All interviews were read by two researchers and the themes were categorized in collaboration.

The groups responded similarly to the questions. All women claimed that they had received sufficient information to decide to participate. All women remembered the invitation and the attached questionnaire, but only a few remembered having received an information leaflet. They remembered information on practical aspects, radiation risks, and arguments for participating. The information that mammography could save lives was considered the most important, and all other information was considered less relevant. The women's primary motivation for participation was to get a confirmation that they did not have breast cancer, which was given as an explanation for why the information was perceived as sufficient. The main argument for participation was therefore not related to the content of the information leaflet. The women wanted more information on the results from the screening program, and also on side effects such as radiation risks and cancer being overlooked, to make informed choices. One participant expressed disappointment when reading about uncertainty of mammography screening.

*Characterization, strengths and limitations*
The authors used qualitative methods including focus group interviews with analysis and discussion of the responses.

The study design is well suited for capturing the main views of the participants and may provide information on how opinions and views on mammography screening are formed.

Participation was low for those who were invited after their first attendance (14/80). The women are not representative of all women who are invited to screening,

since they were included based on their participation. Thus, it would be expected that they considered that they had received sufficient information to choose participation.

*Conclusions*
Women who participated in mammography screening tend not to remember the information leaflet or its content, but still perceived information as sufficient to choose participation.

**Women's experience with mammography screening through six years of participation – A longitudinal qualitative study**
Solbjør M, Skolbekken JA, Østerlie W, Forsmo S
Health Care Women Int. 2014 Dec 15:0.

*Summary of methods and results*
The authors studied how women's experiences with mammography screening developed over time and how long-term screening participation contributed to medicalization. The participants were the same 24 participants as in the book chapter by Solbjør from 2012 [152], discussed above, who were interviewed in 2003 before and after their first NBCSP participation and again in 2009 after three invitations. This study was based on the same interviews and the methods used for analyses of the group discussions were therefore the same.

Attending the screening examination was experienced as being included in a production line, given little personal attention and time for asking questions to the staff. This was uncomfortable at first, but became routine on subsequent screening examinations, and the participants regarded the effectiveness of the system as more important than their individual needs. Experiencing and worrying about pain during the examination was a major topic in 2003, but in 2009 pain was seen as something that had to be endured to have the benefit of screening. The women were less worried while waiting for the results after subsequent screening rounds than after the first screening. Receiving the letter with the result was still a stressful moment, in particular for those who had a previous recall due to a false-positive mammogram. The women perceived the risk of breast cancer in the population in general as high, whereas their individual risk was perceived as low. The main motivation for participation was to get a reassurance that they did not have breast cancer. At the same time, the program in itself contributed to the need for reassurance, since its existence supported the view that a cancer could be present despite feeling healthy.

*Characterization, strengths and limitations*
The authors used qualitative methods including focus group interviews with analysis and discussion of the responses. The study design is well suited for capturing the main views of the participants and may provide information on how mammography screening affects women over a longer time-period.

The participants in this study had attended several mammography screening examinations, and more than half had been to mammography even before the first NBCSP invitation. It is likely that they would be more positive towards screening than the general population of invited women. The fact that two women declined to participate due to emotional distress further supports that there are more and other experiences of mammography than the ones captured by these interviews. Group interviews with women

who already know each other may impose some restrictions on which opinions will be presented.

*Conclusion*
For women who attended several screening rounds, the examination in itself became routine, whereas receiving the result continued to induce emotional stress. This finding may not be representative for the experience of all women who attend screening.

### 5.4.2 Studies outside the evaluation

**The cumulative risk of false-positive results in the Norwegian Breast Cancer Screening Program: Updated results.**
Roman M, Hubbard RA, Sebuodegard S, Miglioretti DL, Castells X, Hofvind S.
Cancer 2013;119:3952-8.

*Summary of methods and results*
The aim of the study was to estimate the cumulative risk of having at least one false-positive result for women who enter the screening program at 50 years and attend 10 screening rounds. Women aged 50 and 51 years at their first screening attendance were included and followed for up to six screening rounds during the period 1996-2010. Women recalled for technical reasons and women with self-reported symptoms were excluded. 231 310 women had one screening round, gradually decreasing to 30 077 women who had six screening rounds.

The cumulative proportion of recalls after false-positive mammograms was estimated for all recall examinations regardless of diagnostic procedures and for recall leading to invasive diagnostic procedures. False-positive results were defined as any recall for further assessment where breast cancer was not diagnosed within 4 months, regardless of the type of diagnostic procedures performed. False-positive results after an invasive procedure included fine needle aspiration cytology, core needle biopsy and/or open biopsy with benign histology. The probability of false-positive results at each examination was estimated with general linear mixed models, with adjustment for year and county. Women who had a false-positive result in a given screening round were excluded from the analyses of risk in subsequent rounds. Probability in the $7^{th}$ to $10^{th}$ round was assumed to be equal to that in the $6^{th}$ round.

The proportion of women with a false-positive mammogram was highest in the prevalence round (5.8%, 95% CI 5.7 to 5.9), particularly the prevalence rounds between 2008 and 2010 (6.9%, 95% CI 6.7 to 7.1). There was a gradual decline towards the $6^{th}$ screening round, when probability was 2.0% (95% CI 1.8 to 2.1). Estimated cumulative probability was 20.0% (95% CI 19.7 to 20.4) for women who enter the program at 50-51 years and attend all 10 screening rounds. Accounting for the possibility of dependent censoring gave very similar results.

The probability of a false-positive mammogram leading to an invasive procedure was also highest in the prevalence round (1.7%, 95% CI 1.6 to 1.7), and cumulative risk for 10 screening rounds was 4.1% (95% CI 3.9 to 4.3). In a given screening round, irregularly attending women had a modestly increased probability of both false-positive results in itself, and of false positive results leading to an invasive procedure.

*Characterization, strengths and limitations*

Women were followed from entering the screening program at the lower age limit using a classical cohort design.

The design chosen will provide an estimate of the proportion of women recalled at least once during 10 screening rounds with the conclusion of a false-positive screening. In a fully implemented screening program, such an estimate may be useful in decision-making for women receiving their first invitation. The individual level of information and details on the diagnostic work-up are additional strengths of the study.

The assumption that the rate of new false-positive tests is constant in the last five screening rounds may be conservative, since the authors observed a decline over the first five rounds. The authors show in a different publication that there is variation between the screening centers [153]. Considering that the number of screening rounds performed by each center varies, the estimate provided must to a large degree depend on the false positive rates in the pilot counties.

At the individual level, the risk of a false-positive result varies with mammographic density [109], which is associated with hormone therapy and other breast cancer risk factors [154]. At a population level, the risk of a false-positive result will depend on the distribution of such factors in the population and on technical and organizational aspects of the screening program that affects sensitivity (digital versus film mammography, double versus single reading, experience/reading volume of the radiologist, acceptance of mistakes and trends in medical ethics – i.e. which types of error are most unacceptable). Variations over time in the factors mentioned above will influence the risk of a false-positive result at a population level. Therefore, the estimate of 20% should not be considered as a fixed number. In particular, the peak in hormone therapy use during the study period, as well as the transition to digital mammography, should be expected to influence the false-positive rates.

*Conclusions*

The authors estimated that among women who enter the screening program when reaching the lower age limit and attend all ten screenings round, 20% will experience at least one recall with the conclusion of a false-positive mammogram and 4% will experience at least one recall with invasive procedures as a result of a false-positive mammogram. These estimates may change over time.

**The cumulative risk of false-positive screening results across screening centers in the Norwegian Breast Cancer Screening Program.**

Roman M, Skaane P, Hofvind S.
Eur J Radiol. 2014 Sep;83(9):1639-44.

*Summary of methods and results*

The aim of the study was to investigate variation in cumulative proportion of false-positive mammograms and in the predictive value of a positive test (PV+) across the 16 screening centers in NBCSP. All attending women 50-69 years during the period 1996-2010 were included (n = 618 636). The definitions of false-positive results were as in Roman et al, 2013 [155] , see above. The statistical methods were also similar, with the use of generalized linear models adjusted for year and age (50-54, 55-59, 60+) to estimate the probability of a false-positive result in each round. Women recalled for technical reasons and women with symptoms were excluded. Women with a false-positive result in

a given round were excluded from the estimation in subsequent rounds. The maximum number of screening rounds was six, and probability in the 7th to 10th round was assumed to be equal to that in the 6th round.

The rate of screening detected cancer varied from 4.2/1000 to 6.4/1000 (mean 5.5/1000). The cumulative probability of a false-positive result varied from 10.7% to 41.5% (mean 23.0%). The cumulative probability of an invasive procedure with benign outcome varied from 2.9% to 12.4% (mean 5.3%). The detection rate and the cumulative probability of a false-positive result were positively associated.

PV+ of recall ranged from 12.0% to 19.9%, whereas PV+ of invasive procedures was much higher and ranged from 28.0% to 58.4%.

*Characterization, strengths and limitations*

Variation in recall rates and PV+ across screening centers was examined using an open cohort design, with individual level information on all variables.

Although variation between centers is an important aspect in evaluation of program quality and the potential for improvement, the results from this study should not be interpreted as reflecting only the quality of the examinations performed at each center. Much of the variation between the counties might be explained by differences in breast cancer risk and breast cancer risk factors, as well as organizing of the screening services. For example, the incidence was higher in the pilot counties compared to the other counties even before screening was implemented. A higher detection in counties with a higher underlying incidence must be expected. Throughout the implementation period, women in the pilot counties also had a higher use of hormone therapy, which increases the risk of both breast cancer and false positive and false negative results.

Since the risk of a false positive mammogram decreased with increasing number of screening rounds [155], the different number of screening rounds in each county may contribute to the variation. No information is provided to indicate how the variation in number of screening rounds conducted by each center was accounted for in the analyses. Variation in opportunistic screening between counties will also contribute to the variation in detection rates, since extensive opportunistic screening will lead to “depletion” of undetected cancers in the population.

*Conclusions*

A considerable variation in the probability of recall mammography and invasive procedures was observed between the screening centers. This may reflect differences in the number of screening rounds conducted in each center, in the underlying breast cancer risk and in screening performance.

**Health-related quality of life, anxiety and depression related to mammography screening in Norway.**

Hafslund B, Espehaug B, Nortvedt MW.
J Clin Nurs. 2012 Nov;21(21-22):3223-34.

*Summary of methods and results*

The authors investigated health-related quality of life, anxiety and depression before mammography screening compared to a reference population. Women invited to the NBCSP in Sogn og Fjordane and Hordaland counties in January-March 2007 received questionnaires and invitation to the study along with the mammography invitation. 10 017

women were invited, 7801 attended screening and 4249 of these (54%) filled in the questionnaire. The questionnaire contained questions on background demographics, anxiety and depression (Hospital Anxiety and Depression Scale, HADS) and health related quality of life (Short-Form 36, SF-36). These women were compared to a reference population from Statistics Norway of 943 women in the same age group. The reference population was a random sample of the Norwegian population who answered completed the SF-36 questionnaire in 2002. Quality of life scores were compared between the groups using linear regression, with adjustment for age, education level, occupation, number of children and smoking status. There was no comparison group for anxiety and depression, but mean scores were compared to those reported in another population-based study from Norway (the Nord-Trøndelag Health Study, HUNT-2) [156].

The responding screening attendees had less education, but more often senior positions in their occupation, and were more often smokers, than women in the comparison group. The screening attendees reported a higher quality of life than the comparison group for all measures except vitality (energy and happiness), also after adjustments. Mean score for anxiety and for depression was lower than for women in the HUNT-2 study.

*Characterization, strengths and limitations*

The authors compared the health-related quality of life, anxiety and depression level in two cross-sectional studies of women in the same age range.

The results on quality of life are adjusted for several potential extraneous differences between the comparison groups. Still, comparing women from Sogn og Fjordane and Hordaland to women from all counties in Norway, there may be differences that are not accounted for. Women in Sogn og Fjordane have a lower all-cause mortality than women in all other counties in Norway [157] and it may therefore not be surprising that they report a better health related quality of life. One would also expect better health among attendees than among non-attendees (self-selection), and the results may therefore apply only to women attending screening and not to all women receiving an invitation. The estimates for anxiety and depression were unadjusted, and therefore likely to be biased.

*Conclusions*

Women who attended mammography screening in Sogn og Fjordane and Hordaland reported higher health-related quality of life and lower levels of anxiety and depression than a general sample of Norwegian women prior to the mammography examination.

**Effects of false-positive results in a breast screening program on anxiety, depression and health-related quality of life.**

Hafslund B, Espehaug B, Nortvedt MW.
Cancer Nurs. 2012 Sep-Oct;35(5):E26-34.

*Summary of methods and results*

The authors examined changes in anxiety, depression and health-related quality of life for women with a false-positive screening mammography, compared to a group of women with a negative screening mammography. All women invited for screening in Hordaland and Sogn og Fjordane counties from 2007 until early 2008 were invited to complete a

questionnaire before screening. The women who were recalled (n=246) answered the same questionnaire before recall examination (sent with the recall letter), three and six months after recall. Both women with suspect findings and women recalled for technical reasons were included. A sample of 229 women with negative screening mammograms, matched by age and geographical location, were invited to answer the questionnaire again six months after screening. Of the 246 recalled women, only 128 with false-positive results replied. 13 women diagnosed with cancer were excluded. Drop-out continued at each point of measurement, with 77 women completing the questionnaire at six months. 195 of the 229 women with negative results responded to the questionnaire at six months. Anxiety and depression was measured using the Hospital Anxiety and Depression Scale (HADS), and health-related quality of life was measured using Short-Form 36 (SF-36). 4249 women participated. Missing items in the HADS questionnaires were imputed using the mean of the completed item scores for depression and anxiety, respectively. For SF-36, the imputation guidelines developed by the researchers who developed the questionnaire was used. Changes in anxiety, depression and health over time within and between the groups were assessed in linear mixed-effects models, adjusting for education, occupation, smoking, age and number of children as fixed factors.

For women with false-positive screening results, anxiety and depression scores increased modestly from before screening to the time of the recall examination. Anxiety scores dropped during the six months after recall and reached approximately the same level as before screening, whereas depression scores increased further from recall to six months after screening.

No clear differences in general health were found at any time point, but mental health was poorer at recall than before screening.

The women who were recalled had similar scores for health-related quality of life, anxiety and depression as women in the comparison group (who had negative screening mammograms). The screening negative comparison group experienced no substantial changes in any of the outcomes during the six months. At six months, the false-positives reported higher depressions scores and poorer general and mental health than the negative comparison group.

*Characterization, strengths and limitations*

The authors conducted a longitudinal study with repeated measurements of emotional distress following recall after mammography screening.

The main strengths of the study include the use of well-established questionnaires for measurement of health-related quality of life, anxiety and depression, the inclusion of a screening negative comparison group, and accounting for the dependence between observations at different time points.

Women recalled for technical reasons and thus not truly false positives, were also included in the study. If recall for technical reasons induces less emotional distress than recall due to abnormal findings, this could have led to underestimation of the emotional distress among those who have a false-positive mammogram. The large number of drop-outs in this study between each time point of measurement may have resulted in a non-representative study population. Precision was low due to the relatively low number of participants at baseline and large number of drop-outs.

*Conclusions*
Women that were recalled after screening mammography experienced an increased level of emotional distress, which decreased over a period of six month, but did not reach prescreening levels.

**Recall mammography and psychological distress.**
Schou Bredal I, Kåresen R, Skaane P, Engelstad KS, Ekeberg Ø.
Eur J Cancer. 2013 Mar;49(4):805-11.

*Summary of methods and results*
The authors compared psychological distress before and after being declared healthy in women recalled for further investigation in the NBCSP. 640 women recalled after screening mammography at Ullevål hospital in 2009 and 2010 were followed for 4 weeks. The women answered a questionnaire before the recall examination and 4 weeks after receiving the results (526 of the 640 included women answered at four weeks). Emotional distress was measured using the Hospital Anxiety and Depression Scale (HADS). Other characteristics such as education, occupation, previous anxiety or depression, degree of optimism, satisfaction with the information given were also recorded. They were also asked if they would attend again and if they would recommend screening to other women.

The authors compared the level of anxiety and depression at recall and four weeks after the results according to type of recall examination (imaging only, imaging + biopsy) and result (no cancer, cancer). The outcome was measured as a proportion using a cut-off of 11 points in HADS for clinical anxiety and clinical depression, as well as mean HADS score within each group. These comparisons were made using Fisher exact tests/chi-square tests, and t-tests/Mann–Whitney-U tests, respectively.

Among the women recalled, 12.5% were diagnosed with cancer (invasive or DCIS) and the remaining were considered healthy. Thirty percent of the 640 included women had been recalled also at a previous screening round. For women who turned out to have a false positive mammogram (n=560), those who responded at four weeks (n=454) had lower anxiety levels and similar depression levels at four weeks compared to the whole study sample (n=560) at recall (anxiety mean score 6.1 -> 4.5, depression mean score 2.4 -> 2.6). No clear differences could be detected depending on type of recall examination. For women who were diagnosed with cancer (n=80), those who responded at four weeks (n=72) had slightly lower levels of anxiety and higher depression levels at 4 weeks compared to the whole study sample (n=80) at recall (anxiety mean score 6.6 -> 5.6, depression mean score 2.7 -> 3.4). Results were similar for mean HADS score and for proportions above the clinical cut-off values. Nearly all participants stated that they were satisfied that they had participated in the screening program and would attend the next screening round. The strongest predictors of HADS score at four weeks were baseline HADS score and degree of optimism.

*Characterization, strengths and limitations*
The authors conducted a longitudinal study of women recalled after screening mammography.

They used a well-established and validated screening tool for anxiety and depression and were able to investigate psychological distress according to type of recall assessment and outcome.

Possible sources of systematic error in the study may be the 18% drop-outs and the fact that comparison was not restricted to those who answered both questionnaires. Still, it seems unlikely that the observed drop in anxiety levels should be due to sample differences entirely. The study cannot provide information on whether the recalled women go back to the "before-screening" anxiety levels, since this was not measured. However, the mean value of anxiety levels at four weeks is quite similar to those reported for a general population of similar age, whereas mean depression score is somewhat lower [158]. The cut-off value for clinically relevant anxiety or depression in the study was a score of 11, which is higher than the recommended value of 8 [159].

The dependency of the data was taken into account in the analyses, since the tests used for comparison were for independent observations. This will affect the power to detect differences, but not the proportions and mean scores in themselves. The precision of the estimates could not be evaluated since only p-values above or below 0.05 or 0.01 was provided.

*Conclusions*

Symptoms of anxiety declined during the first four weeks following a false-positive recall examination, regardless of the type of investigation performed. Symptoms of depression remained unchanged. For women with breast cancer, anxiety levels dropped modestly over four weeks, whereas depression increased. Changes from before screening and long-term psychological distress could not be addressed in this study.

**Experiences of recall after mammography screening--a qualitative study.**

Solbjør M, Forsmo S, Skolbekken JA, Sætnan AR.
Health Care Women Int. 2011 Nov;32(11):1009-27.

*Summary of methods and results*

The authors examined the experiences and attitudes towards mammography screening among women who had been recalled for further examination due to abnormal findings. Semi-structured, individual interviews were conducted in 2004 and 2005 of women living in Central Norway. The invitation criteria were no self-detected symptoms and living within 45 minutes' drive from the hospital. Of 35 eligible women, eight accepted. The first interviews took place between receiving the recall letter and the follow-up examinations and focused on the experience of receiving a recall letter and awaiting the follow-up examination. The second interview was after the follow-up (immediately after or when they had received the test results) and focused on the experience of the follow-up examination, the test results and attitudes towards mammography screening. At the time of the second interview, two women had been diagnosed with breast cancer following the recall examination, one was waiting for additional examinations, and the remaining knew that the initial mammogram was false-positive. Interviews were audiotaped and transcribed. One author read all interviews and all authors read some interviews. Themes were identified beforehand and during reading and discussions.

The women had received information before the mammography examination that the recall risk was about 3%, but were not prepared to be recalled. The recall letter stated that 20% of the recalled women would be diagnosed with breast cancer after the diagnostic work-up. Some considered this a low number and were reassured, whereas others considered the number high and felt increasingly worried. All women were satisfied that the waiting period from receiving the recall letter to the date of examination

was short (4-5 days), but some also saw this as an indication of the severity of the situation. They also worried about how much a potential cancer could grow from the initial mammography to the recall examination (up to 5-6 weeks).

During the second interview, those who turned out to be false-positives still had worries about the certainty of the findings and whether there was a need for further follow-up. The three women who had been diagnosed with cancer or awaited additional tests expressed both fear and relief to be diagnosed at an early stage. Most remained positive towards the screening program, whereas one woman was convinced that the recall was unnecessary and expressed doubts about continued participation.

*Characterization, strengths and limitations*

The authors used qualitative methods to investigate the experiences of women who were recalled after mammography screening. Individual interviews may be able to capture and display more nuances in the reactions among women recalled after mammography, compared to the quantitative measurements described above [160, 161]. The study provides valuable information for those who administer the recall process and could contribute to improved adaption to these women's needs. Eight women may be too few if the aim is to reflect a broad range of reactions. The low participation may also indicate that there could be other or stronger reactions among those who did not consent. The authors also express concern that the age differences between the interviewer and the participants may have influenced the interview situation.

*Conclusions*

The women were not prepared to be recalled after the mammography screening, despite being aware of the possibility of recall. The information in the recall letter was reassuring to some, and increased worries among others. Those who had false-positive results were not completely reassured by the diagnostic work-up.

**Challenges of informed choice in organised screening.**

Østerlie W, Solbjør M, Skolbekken JA, Hofvind S, Saetnan AR, Forsmo S.
J Med Ethics. 2008 Sep;34(9):e5.

*Summary of methods and results*

The authors investigated experiences related to invitation and decision of participation among women who received their first invitation to the NBCSP. They conducted focus group interviews with women in Nord- and Sør-Trøndelag who had received their first invitation from the NBCSP, and had decided to attend, but not yet attended. Sixty-nine women were interviewed in eight focus groups. Some women who accepted had to be excluded to keep the groups small and to avoid having relatives and neighbors in the same group if possible.

The groups were asked the following questions:

1) Could you describe your thoughts and reflections concerning the mammography you are going to in a couple of days?
2) What thoughts have you had concerning your (own) risk for breast cancer?
3) What thoughts have you had concerning breast cancer prevention?
4) What implications do you think screening programs like mammography have for your health?
5) Would you recommend mammography to other women?

Interviews were audiotaped and transcribed. Coding categories were made in consensus by all authors after reading transcriptions. Two authors analyzed all the material.

The majority had previously had mammography examination outside the program. The women expressed great trust in mammography as a diagnostic tool, and considered a normal mammogram as a proof of healthy breasts. They saw the invitation as a help to get the mammography that they ought to go through anyway in order to take care of their health. They did not see the letter as an invitation, but rather as a call, signaling that others took responsibility. Decision-making was not considered necessary.

*Characterization, strengths and limitations*

A qualitative design with focus group interviews was used to examine the process of deciding to participate in mammography screening.

The study provides information on how invitation for mammography screening is perceived by women who decide to attend, and also on how opinions on mammography are formed in groups of women.

The exact number of invited women was not stated and participation and potential selection of participants is difficult to evaluate. All the interviewed women had decided to attend screening before the interview, and many had demonstrated positive attitudes towards mammography through previous mammography examinations. Thus, their views may not be representative of all women attending mammography and in particular not representative of those who choose not to attend.

*Conclusions*

For the women who participated in this study, attending screening was regarded as obvious due to great trust in mammography. The women in this study may have been more positive towards mammography screening than women in general.

**"You have to have trust in those pictures." A perspective on women's experiences of mammography screening.**

Solbjør M.

In J. Brownlie, A. Greene, & A. Howson (Eds.), Researching trust and health. London: Routledge, 2008.

*Summary of methods and results*

The authors examined women's attitudes towards mammography screening in focus group interviews of women who attended screening. They were invited before, shortly after and six months after the mammography examination took place. Sixty-nine women participated in eight groups (the same study as Østerlie et al [162]). The groups were organized according to age (50-59 and 60-69 years). Some women in some of the groups knew each other. The study was based on the first interview in each group. The questions asked were not directly about trust, but rather about the women's experience in general.

Trust became an explicit discussion theme in six of the eight groups. The women were aware that mammography had some limitations, and could not be fully trusted. The imperfection was perceived as mostly due to technical aspects (will some parts of the breast be missed?), but also to the radiologists' interpretation of the images. Still, they had greater confidence in mammography due to the visualization of the breast than they had in their own or a doctor's palpation of the breast. Their worries were primarily about false negatives.

For the remaining two groups, trust did not become a topic. These groups focused mainly on recent or previous disease among themselves or in their families. The author discuss briefly that this could be due to lack of awareness or that trust was so obvious that they did not feel the need to discuss this.

*Characterization, strengths and limitations*
The authors used a qualitative study design with group interviews followed by analysis and interpretation of the group discussions.

The study design is well suited for capturing the most common views on mammography among screening attendants, and contributes to an improved understanding of their motivation for attending.

The participants in the study were those who had decided to attend screening. It is therefore not surprising that they trusted the screening program and had mainly positive attitudes towards mammography. In addition, being willing to participate in a research project may require a certain level of trust in general. The small proportion interviewed of the total number invited for the study also increases the possibility that there are beliefs and views on mammography screening that were not captured in this study, in particular among those not attending screening.

*Conclusions*
The women who participated in this study expressed great trust in mammography screening due to the visual nature of the examination. The possibility of a false negative result was the primary source of distrust. Women who do not attend screening may have different opinions.

### 5.4.3 Summary

The included studies provide information on the experiences of women who attend screening without being diagnosed with breast cancer, neither at screening nor during screening intervals. In addition, two studies provided information on the experiences of women recalled for further diagnostic tests resulting in a breast cancer diagnosis.

Two studies provide estimates on the frequency of being recalled for further examinations due to abnormal mammograms, with subsequent normal findings. The cumulative probability of a first-time false positive mammogram for women who attend all 10 screening rounds was estimated to approximately 20%. Since information from 10 screening rounds is not yet available, the authors made assumptions that the proportions in the last rounds would be similar [155]. The probabilities of false positive mammograms differed between the screening centers/counties, a finding that may reflect differences in quality, but also in breast cancer risk and distribution of risk factors, as well as differences in the number of screening rounds.

Two quantitative studies present results on the mental distress associated with recall [160, 161]. The studies indicate that mental distress is high at recall, but declines over time. For women who were diagnosed with breast cancer during recall examinations, depressive symptoms increased during the period following recall examination. The number of participants in these studies is small, and in addition they both suffer from a relatively high number of drop-outs, increasing the risk of selection bias. Individual interviews of women recalled after mammography screening supports the findings of increased mental distress due to recall, and indicate that women who were recalled were

not prepared for this, despite receiving information on the possibility of recall before screening [163]. The information in the recall letter was perceived by some as reassuring, but as worrying by others. They were content that the time from the recall letter to the recall examination was short, although some also saw this as a sign of severity. Most of the women remained positive towards the screening program.

In one study, the mental distress among women invited for mammography was investigated [164]. Due to the study design, only those who decided to attend screening participated in the study. These women had a lower level of anxiety and depression than the general population, which may be due to selection of particularly healthy participants.

Qualitative studies based on focus group interviews of women attending screening [152, 162, 165, 166] indicate that women who attend screening are aware that cancers may be missed, but still trust the screening program more than they trust clinical examination/self-examination. They do not see the invitation as an invitation, but rather as a call, rendering decision-making redundant. Most women could not remember that they had received an information leaflet, but still felt that they had sufficient knowledge to attend. The primary aim for those who attend is to get a confirmation that they do not have breast cancer. Continued participation increases the feeling of routine and being part of a production line. Pain and emotional distress is perceived as less dominating in subsequent screening rounds compared to the first participation.

## 5.5 Costs and cost-effectiveness

### 5.5.1 Studies in the evaluation

**Direct and indirect costs of the Norwegian Breast Cancer Screening Program**
Moger TA, Kristiansen IS
University of Oslo, Health Economics Research Programme, Working paper 2012: 3
Note: The steering committee also received a revised version of the published paper. To our knowledge, the revised version has not been published.

The authors estimate the total costs of one screening round among women aged 50-69 years including costs of recall examinations in the NBCSP. Information on attendance per county was obtained from the Cancer Registry of Norway, and a recall rate of 3.5% was assumed based on a previous publication [167]. In NBCSP a total of 27 screening units at hospitals and four mobile mammography units (buses) are operating. The authors assessed the direct health care costs, travel costs as well as productivity loss associated with women attending screening and recall examinations. Screening costs included capital and operating costs of mammographs and mammography buses, costs of office space and personnel (radiologists and technicians), and costs of postal letters (invitation, reminders, examination results, etc.).

For recall examinations, the authors assume that all women have a new mammogram of one breast, and that 50% have an ultrasound examination and a biopsy. The unit costs of clinical mammography and ultrasound were estimated using reimbursement rates for these procedures and out-of-pocket payment rates for 2012. The costs of biopsy were derived by means of a cost model. Health care costs were obtained from public data (fee schedules/reimbursements) and from personal communication with

employees at the Oslo University Hospital, University Hospital of North Norway and Nordland Hospital Trust.

Travel costs were estimated at the municipality level, using the mean travel distance to the nearest mammography unit multiplied by the government refund fee per kilometer for travelling by car, NOK 3.90. Maximum travel costs per visit were set at NOK 1500. Productivity loss was estimated using the national mean pre-tax salary (including social costs) for women 50 years and older, the proportion of women 50-69 years working in each municipality, the mean round-trip travel time to the nearest mammography unit, and an examination time of 0.25 hours for screening and 0.5 hours for recall examination. Travel time was longer for recall examinations since these in many cases would be performed at a more distant hospital. The authors assume that 20% of the screenings takes place at the mobile screening units.

The 2012 costs of one screening round excluding recall examinations were estimated to 521 million NOK, corresponding to an average of 1262 NOK per woman examined (SD 353 NOK). Health care costs constituted 64% of the total costs, whereas travel costs and productivity loss constituted 15% and 21% of the total costs, respectively. Repeat examinations involve more medical procedures and thus also higher direct health care costs than do screening. Total costs of recall examinations in one screening round were estimated to 52.8 million NOK, of which 79% were direct health care costs, 9% travel costs and 12 % productivity costs. The average costs of per women attending a recall examination were estimated to 3655 NOK (SD 761 NOK). The total costs of one screening round in 2012, including repeat examinations, were estimated to 574 million NOK, in average 1389 NOK per woman attending screening. Referring to an article by Hofvind in 2007 [167] the authors assumed that 15.9% of the repeat examinations are true positive (cancer) and assessed the costs of false positive mammograms to a total of 44.4 million NOK per screening round.

*Characterization, strengths and limitations*

Inclusion of travelling costs and productivity loss contributes to a more complete estimation of the costs for society as a whole, and represents the main strength of the study. The study indicates that travel distance and travel time contribute considerably to the costs of screening.

The main limitation of the study is the generalization of the cost estimates. Data has not been collected from individual mammography units or diagnostic centers. The authors estimate the costs of one screening round based on a simple model of the different cost components and their estimated unit costs. There is no explicit discussion of the assumptions underlying their model. The unit price of the different cost components are set mainly based on personal communication and on guesstimates, however public reimbursement rates and out-of pocket payments are used when deriving the costs of clinical mammography and ultrasound.

Due to lack of information, the authors used national level figures for parameters that may vary across counties and municipalities. Recall rates vary greatly across the mammography screening centers in Norway [153], and the assumption of 3.5% recall rate for all units may also give an inaccurate estimate of costs. It is expected that a 1% increase/decrease in the re-examination rate will increase/decrease the costs of recall examinations by around 28% (Moger, TA, personal communication December 1st 2014). In 2012-2013, the average recall rate was 2.8% for all counties combined, and 18.9% of the recalled women were diagnosed with breast cancer in recall examinations. The

authors used NBCSP attendance rate per county in 2008 since information at municipality level was not available.

NBCSP costs not directly associated with the screening or recall examination itself are not included in the study. Among these are costs associated with administrating and operating the screening program as such, e.g. administrative costs, IT-systems, coding and registration, quality assurance, and training of staff, etc. In addition, and as the authors state in their paper, the costs for repeat examinations do not include costs outside the mammography screening units, such as visits to the physician and surgical conferences etc.

*Conclusions*

The total costs of one screening round were estimated to 574 million NOK, amounting to an average of 1389 NOK per woman attending screening in 2012. Travelling costs and productivity loss constitute a considerable part of these costs. The cost estimates are based on fairly crude aggregated data. Cost associated with running the NBCSP program as such are not included in this study, and the total costs of the NBCSP are therefore likely to be underestimated.

**Expected ten year treatment cost of breast cancer detected within and outside a public screening programme**

Moger TA, Bjørnelv GMW, Aas E
Submitted manuscript

*Summary of methods and results*

The authors estimated ten-year expected treatment costs for breast cancer among women aged 50-69 years according to detection mode and stage at diagnosis. All information was at an individual level and was obtained from data linkage between the Norwegian Patient Registry, the Norwegian Prescription Database, the Cancer Registry of Norway and Statistics Norway. Costs of hospital treatment were studied for women diagnosed with breast cancer between 1999 and 2009. Information on hospital treatment prescribed hormonal treatment was available for 2008 and 2009. Costs for the different types of treatment were estimated by the adjusted DRG weights from the Patient Register, with addition of patient co-payments for out-patient visits. Costs for hormonal treatment were based on retail sales prices. Other drug treatments, such as chemotherapy were considered to be included in the DRG weights. Treatment costs were allocated to two-month periods since time of diagnosis. The proportion receiving each treatment was calculated based on the number of women who were in a given two-month period since diagnosis during 2008 and 2009. This proportion was used to calculate mean treatment costs for each two-month period. Cox proportional hazards models with all-cause mortality were used to estimate the proportion of patients alive in each two-month period and thus at risk of receiving a given treatment. Treatment costs per two-month period were calculated as the mean treatment cost multiplied by the proportion of patients alive during that period. A discount rate of 4% per year was used. Treatment costs were compared for screening-detected cancers, interval cancers, and non-participants. The latter group included both women not yet invited and women invited but not attending.

The estimated mean expected ten-year treatment costs (measured in 2008 prices) for women with breast cancer regardless of detection mode and stage were € 44 490 (95%

CI € 43 070 to 46 020), with lower costs when screening-detected than if detected outside the NBCSP (referred to as non-attenders): € 43 240 and € 49 670, respectively (95% CI for difference € 3040 to 10 230). Estimated mean expected ten year treatment costs for interval cancers were € 61 610. Expenditure on chemotherapy was higher for interval cancer and for non-attending women than for screening-detected cancer, whereas expenditure on radiation therapy was higher for screening-detected cancers. Treatment costs increased with advancing stage at diagnosis from € 21 990 (95% CI € 19 370 to 24 640) for DCIS to € 83 700 (95% CI € 71 100 to 98 950) for TNM IV. In the early stages, screening-detected cancers had lower treatment costs than cancer detected outside the program, while costs were higher at more advanced stages.

*Characterization, strengths and limitations*

The authors combined data on resource use associated with treatment of breast cancer in hospitals over a two-year period with data on survival experience for breast cancer patients over a ten-year period, and estimate the expected ten-year treatment costs of breast cancer detected within and outside the NBCSP. The design allows estimation of long-term treatment costs despite the fact that information on treatment and the corresponding costs was only available for the latest part of the study period.

The observed cost window (2008-2009) allows for estimating costs for updated treatment practice. However, the number of patients eligible for receiving the different treatments in each two-month period will be small and as a result the uncertainty of the estimates will be high. Changes in treatment during the study period could affect the proportion of patients alive in each two-year period and consequently expected treatment costs.

The analysis encompasses costs of hospital treatment and prescribed hormonal treatment. Costs associated with other types of health care services are not included. The estimates of expected treatment costs are based on DRG weights for the different procedures and DRG unit price as defined in 2008 and 2009. These are regularly updated and will hence affect the estimates of treatment costs.

Treatment costs for women with breast cancer falling into the group termed 'non-participants' can differ between the not-yet-invited and the true not-attending.

*Conclusions*

Overall, the expected ten-year treatment costs for breast cancer based on treatment in 2008-2009 were € 44 490 (95% CI € 43 070 to 46 020), corresponding to NOK 356 000 (95% CI NOK 345 000 to 368 000), all measured in 2008 prices. Costs increased according to stage at diagnosis, and were higher for non-participants and interval cancers compared to NBCSP attenders.

**Research-based evaluation of the Norwegian mammography screening programme; effectiveness, side-effects and cost-effectiveness**

Van Luijt PA, Heijnsdijk EAM, de Koning HJ

Final project report to the Research Council of Norway, 2015

*Summary of methods and results*

The authors used the MIcro-simulation SCreening ANalysis (MISCAN) model, developed in the 1980-ies, to model the expected trends in incidence and mortality in Norway following screening implementation under various assumptions. The validity of

the different models was assessed through comparison with the observed trends. The same models were used to estimate change in breast cancer mortality, level of overdiagnosis, and cost-effectiveness. A more detailed description of the simulation models were given in section 5.1.1.

Three models with different assumptions were tested:

Model 1: Simulating NBCSP screening and opportunistic screening (using data from the publication by Lynge et al [117]) as well as a risk ratio of 2.2 for women using hormone therapy, using sales numbers and summary data from the Norwegian Prescription Database to model the extent of use, by age and year. Hormone therapy was assumed to increase the onset of disease (i.e. the rate of DCIS), but not disease progression (i.e. transition times).

Model 2: In addition to Model 1, another risk factor was added, increasing the number of women developing breast cancer. All women aged 87 and younger in 1997 were modeled to have an additional risk factor for breast cancer in the years 1997-2006 that increased the age-specific hazard with a factor of 1.75.

Model 3: Similar to Model 1, but with an additional assumption of very slow growing tumors, i.e. a large pool of dormant disease. This was achieved by allowing the model to use wider boundaries for the dwell time parameters. For example, dwell time for DCIS was 0.4 years in Model 1 and 4.79 years in Model 3.

According to the authors, Model 2 provided the best fit for incidence and mortality combined. Model 1 gave a poor fit both for incidence and mortality; whereas model 3 resulted in a model fit more comparable to that of model 2. For Model 2, the authors also conducted sensitivity analyses using different levels of screening sensitivity (50 and 100%), attendance (50 and 100%) and screening intervals (1 and 4 years). In these sensitivity analyses, opportunistic screening was not included. For cost-effectiveness, an additional sensitivity analysis with high and low positive predicted value was conducted.

Information on costs of screening and of treatment was obtained from the studies by Moger et al (summarized above). Ten-year treatment costs according to stage at diagnosis were applied to the predicted number of cases in each stage with and without screening in the different models. Cost-effectiveness was estimated as costs per quality adjusted life year (QALY) gained, compared to a situation without screening. QALYs were calculated from the utilities described by Haes et al [168] and information on treatment and disease-free survival according to disease stage at diagnosis from the Cancer Registry of Norway.

Cost-effectiveness was estimated for women born in 1955 and followed throughout 2055 who had a predicted reduction in breast cancer mortality of 16% in Model 1 and 2 and 13% in Model 3, respectively. Effects and costs were discounted at 3.5% per year. For all analyses, a population of 10 million women was simulated.

The direct costs per QALY gained were estimated to be NOK 127 317, NOK 112 162 and NOK 302 315 in Model 1, 2 and 3, respectively. When including also indirect costs, the estimates increased to NOK 211 525, NOK 189 557 and NOK 478 576. In sensitivity analyses using Model 2, the estimates for total costs (direct and indirect) per QALY varied between NOK 158 000 and NOK 261 000. The screening parameter with the greatest impact on costs per QALY was variation in the screening interval.

*Characterization, strengths and limitations*

The MISCAN model is developed especially to study screening outcomes and has well-known characteristics as it has been tested in several other settings. The modeling could use empirical Norwegian data for key variables. A model study can test if several different theoretical scenarios are reasonable, i.e. compatible with the observed incidence and mortality trends. Since the MISCAN model allows simultaneous estimation of both benefits and harms of screening, it provides a method for taking most health consequences of screening into account in the cost-effectiveness balance.

A limitation of all modeling studies is that the results are to a varying degree sensitive to assumptions about phenomena for which there are no empirical data, either because of missing information or because they in essence are not directly observable. The failure of Model 1 to predict the marked increase in incidence from 1994 may in part be due to the association between hormone therapy and screening attendance (both program and non-program). Furthermore, the differences between counties in prescreening incidence and hormone therapy use throughout the period may also play a role, while the models were calibrated using national level data. We are not aware of any single risk factor that could fit the characteristics for the additional risk factor in Model 2, and smaller increases in several risk factors may be a more plausible explanation.

The information on costs of screening and breast cancer treatment was obtained from the studies by Moger et al, and the limitations of those studies will thus also apply to the estimates of cost-effectiveness. Cost of screening are measured in 2012 NOK while cost of treatment is measured in 2008 NOK and increase in prices during this time interval is not accounted for. The cost of treatment for cancer among women older than 69 years could be different from those estimated by Moger et al, for example due to more co-morbidity among older patients. If so, the cost reductions from earlier detection would be expected to be larger than estimated in this study.

The utilities used to calculate QALYs were obtained from a study based on breast cancer treatment in the 1980s. Quality of life associated with different treatment modalities may have changed considerably since then. It is important to note that if life years gained had been used instead, the calculated estimates would be more cost-effective.

It should be noted that since the cost-effectiveness estimates were calculated for a birth cohort followed from their first invitation at age 50 years and throughout life, the estimates of effectiveness and overdiagnosis are not directly comparable to those predicted for the population of invited women until 2025. The estimated levels of overdiagnosis in the MISCAN model are modest compared to the studies of overdiagnosis in the NBCSP that are based on observed data only. A high level of overdiagnosis implies higher costs (both due to recalls and possible unnecessary treatment), and also reduced quality of life during the relevant time period. Hence, the cost-effectiveness ratio would be higher relative to a situation with a low level of overdiagnosis.

*Conclusions*

The study indicates that the NBCSP is highly cost-effective at modest levels of overdiagnosis. The validity of the estimated cost-effectiveness depends on the validity of a range of factors that are not easily assessed, and the uncertainty of the estimates is therefore considerable.

### 5.5.2 Studies outside the evaluation

We have not found any publications on costs or cost-effectiveness of the NBCSP from 2008 or later.

### 5.5.3 Summary

The studies by Moger et al estimate the costs associated with breast cancer screening and treatment. The social costs of one screening round was estimated to NOK 574 million, or NOK 1389 per woman attending screening in 2012, including costs of recalls after positive mammography. The estimated 10-year treatment costs for breast cancer were NOK 356 000 (95% CI NOK 345 000 to 368 000) measured in 2008 prices. The costs to society of NBCSP may be underestimated, since not all relevant costs were included.

The results from the two papers by Moger et al were used in the MISCAN simulation model to estimate the program's cost-effectiveness. The estimated cost-effectiveness ratio is dependent not only on the available data on cost of the NBCSP, but also on the estimated effectiveness of the program. The screening program's effectiveness is depending on breast cancer incidence and the programs impact on breast cancer mortality. Also factors such as screening attendance, screening sensitivity, breast cancer incidence, rate of overdiagnosis and overtreatment, extent of opportunistic screening, cancer treatment are important. In addition, the calculated number of quality-adjusted life-years gained from screening is dependent on the estimated net gains from reduced morbidity and mortality, as well as the utility weights and probabilities assigned to the different phases of the disease (health states). All of these factors affect the estimated cost-efficiency ratio.

Van Luijt et al estimate a cost per quality-adjusted life-year gained (NOK 190 000 to 479 000) which is below the cutoff value recommended by WHO, which for Norway would be approximately NOK 1 900 000. In other settings, the Norwegian Directorate of Health has used cost thresholds of NOK 400 000 – 1 000 000 per life-year gained.

# 6 Discussion

## 6.1 Summary of results

The basis for this evaluation has been observational studies of the Norwegian Breast Cancer Screening Program published from 2008 to 2014, most of which used data from the same quality assured project database. Results within each topic in the evaluation were summarized following assessment of the design and methods used with respect to the validity and precision of the estimates in the individual studies.

All studies providing an estimate of program effectiveness indicated a reduction in breast cancer mortality for women invited to mammography screening, but with variation in the magnitude of the reduction and in statistical precision. Use of individual data on invitation date, diagnosis and death, long follow-up and detailed adjustment for time trends and regional differences were considered as important factors in validity assessment. A summary measure across the studies of reduction in breast cancer mortality attributable to the implementation of the NBCSP, compared to a situation with no screening program and with emphasis on the most reliable estimates, is considered to be in the range 20-30% for women aged 50-79 years.

In studies of overdiagnosis following NBCSP implementation, the variation in study design and analytical approach, and also in the choice of denominator for overdiagnosis, resulted in a large range of estimates. Again, individual data, long follow-up and assumptions on breast cancer incidence in the absence of screening were key factors in estimation of the excess incidence in screening and the post-screening drop in incidence. We consider the most reliable estimates of overdiagnosis of invasive breast cancer and DCIS combined, for women aged 50-79 years compared to a situation without screening, to be within the range 15-25%. For women aged 50-79 years in a situation with screening, we consider the corresponding estimates to be within the range 15-20%. These estimates correspond to the societal perspective on overdiagnosis, as described in sections 4.3.1 and 5.2.3. It should be emphasized that both as a measure of costs (the societal perspective) and as a measure of possible harms (the individual perspective), the summary estimates of overdiagnosis do not reflect the basic expectation that diagnoses in a screening situation will be on average ‘milder’ than those in the absence of screening [54], as described in section 1.2. On average, an overdiagnosed breast cancer would therefore be expected to need less aggressive treatment, cause less harm and generate lower additional costs than would a clinically detected breast cancer [14].

The total costs of one screening round was estimated to NOK 574 million, or NOK 1389 per woman attending screening in 2012, including costs of recall examinations and indirect costs, but not administration costs. The estimated 10-year treatment costs for breast cancer in 2008 were NOK 356 000 for treatment of one patient. These costs were used to estimate cost-effectiveness of NOK 190 000 to 479 000 per quality-adjusted life-year gained by screening, and depended mainly on the number of breast cancer deaths prevented.

Interval cancers were estimated to comprise approximately 25% of cancers among women attending screening, including both invasive cancer and DCIS. One third of interval cancers may be tumors missed at previous screening (false negative), whereas two thirds may be true interval cancers. Diagnostic delay due to negative screening

mammography exists, but the extent is not known. Women with interval cancers generally remain confident in the screening program.

Among women attending all 10 screening invitations, approximately 20% were estimated to experience at least one recall for further examinations due to false positive results. The increase in mental distress following a recall examination declines over time, but may recur at subsequent screening invitations or examinations. Women attending screening express greater concern for interval cancers than for false positive results. The primary motivation for attendance is to get a confirmation that they do not have breast cancer. Continued participation increases the feeling of routine and being part of a production line. Pain and emotional distress is perceived as less dominating in subsequent screening rounds compared to the first participation.

### 6.1.1 Balancing benefits and harms of the Norwegian Breast Cancer Screening Program

To enable a more direct comparison of the results summarized above, we have applied the results in absolute numbers to an expected cohort of 10 000 women aged 50 years who are invited for 10 screening rounds and followed for their remaining lifetime. We assume that 76% of the invited women attend all 10 screening examinations, whereas the remaining 24% never attend. All calculations were made for invasive cancer and DCIS combined. The outcomes of mammography screening for this expected cohort are summarized in the figures below per 10 000 invited women. We emphasize that there is considerable uncertainty in these numbers, reflecting both the uncertainty in the included studies and in the assumptions made in the calculations.

*Breast cancer deaths prevented*
The number of breast cancer deaths prevented in the situation outlined above was by Weedon-Fekjær et al, corresponding to one breast cancer death prevented per 368 women invited (95% CI 266 to 508) or 27 breast cancer deaths prevented per 10 000 invited women. The details of the calculation are available in the study's web appendix [120]. The calculations were based on the observed breast cancer mortality and total mortality in 2009.

*Overdiagnosis*
To obtain comparable absolute estimates of overdiagnoses, we applied incidence rates from the same period (2006-2009) in combination with the total mortality as provided by Weedon-Fekjær et al. Incidence rates for invasive breast cancer and DCIS combined for the period 2006-2009 were provided by Duffy et al for the age categories 50-54, 55-59, 60-64 and 60-69 years [128]. Based on these rates, the expected number of breast cancer diagnoses from age 50 to age 69 in a situation with screening (i.e. the denominator in method C) was 569. The details of the calculation are available in Appendix IV. The estimate of overdiagnosis as measured with method C from this evaluation was 20-30%, which corresponds to approximately 114-170 breast cancer diagnoses.

Method C was chosen in this calculation since nationwide observed data for the denominators in method A and B are not available: The incidence rates among women aged 50-79 years in a situation *without* screening (A) are by definition not observable in a population offered screening. The incidence rates among women aged 50-79 years in 2009 were still partly influenced by women who had not been invited for screening and

did therefore not correspond to a situation *with* screening (B). We note, however, that the absolute number of overdiagnoses as estimated by Sørum Falk et al [83], using the observed incidence rates in 1980-1984, corresponds to 103 (95% CI 63 to 143) overdiagnoses among 7600 regularly attending women[54]. Although lower than our estimate using method C, the large uncertainty in both estimates indicates that they may be considered as comparable.

*Balancing breast cancer deaths prevented and breast cancer overdiagnoses*
Combined, we consider the numbers presented above to be compatible with approximately five overdiagnosed breast cancers per breast cancer death prevented.

*Screening-detected breast cancer*
The expected number of screening-detected breast cancer were calculated using detection rates for the period 2006-2009 obtained from the project report by Duffy and Michalopoulos, for the age categories 50-54, 55-59, 60-64 and 60-69 years [138]. Details of the calculations are available in Appendix IV. The expected number of screening-detected breast cancers was 377.

*False positive and false negative mammograms*
The absolute number of women who experience at least one recall examination with the conclusion of a false positive mammogram was estimated to 20% of screening attendants by Roman et al [155], corresponding to 1520 women among 7600 attendants. The majority of these women will be cleared from cancer suspicion after a second mammography or an ultrasound examination, whereas 4.1% or 310 attending women would be expected to be cleared for cancer suspicion only after invasive tests such as cytology or biopsy.

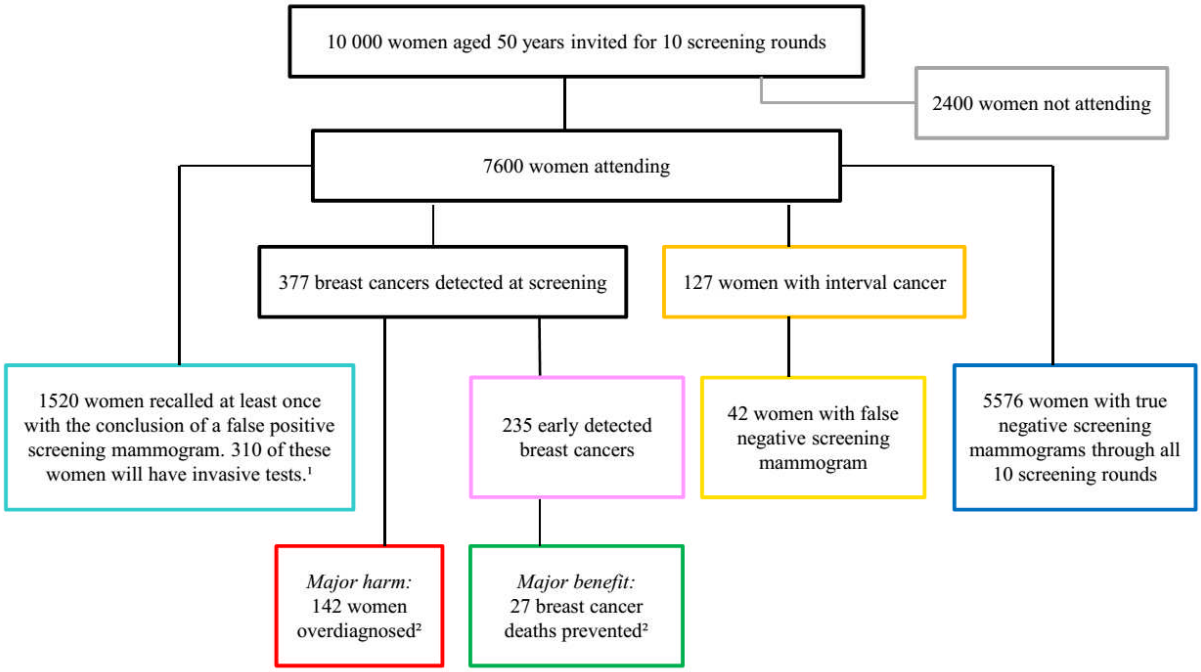
A proportion of the women who have a false positive mammogram will subsequently have a breast cancer detected at screening or during screening intervals. Due to lack of information on the risk of breast cancer for women with a false positive mammogram in the NBCSP, we could not account for this in the summary figures below.

In calculations of the expected number of interval cancers, we used descriptive data presented in the project report by Duffy and Michalopoulos [138] and in Hofvind et al [151]. These studies were chosen since they provided the most updated numbers and included nationwide or close to nationwide data, respectively. We calculated the expected number of interval cancers diagnosed following 10 screening rounds assuming an interval cancer rate of 0.55/1000 screens during the first year and 1.18/1000 screens during the second year (combined 1.73/1000 screens) of the screening interval. Due to lack of age-specific data, we assumed common rates of interval cancer across all ages, which may not be realistic. Details of the calculations are available in Appendix IV. The expected number of interval cancers was 127. Assuming that 30-35% of these had a false negative screening mammogram at their previous screening, we expect that approximately 42 would have a false negative mammogram.

[54] Sørum Falk et al 2013 estimate 428.4 overdiagnoses among 31 586 attending women, corresponding to 19.4% overdiagnosis using method A (95% CI 11.8 to 27.0%). For 7600 attending women, this corresponds to
(7600 / 31 586) * 428.4 = 103 overdiagnoses.

*True negative mammograms*

The remaining 5576 of the screening attendants would be expected to have only true negative mammograms at all 10 screening rounds. This number should be considered as a conservative estimate (i.e. the true number should be expected to be higher) since some of the 1520 women with false positive screening mammograms would subsequently have a breast cancer detected at screening or during screening intervals and therefore also be included among the 377 women with screening-detected breast cancer or the 127 women with interval cancers. Unfortunately, we did not have sufficient information to estimate the number of breast cancers detected during the screening age range among women with false positive results.




[1] Some of these women may be diagnosed with breast cancer at subsequent screening rounds or during a screening interval. [2] When followed for their remaining lifetime

*Note: There is considerable uncertainty in these numbers, reflecting both the uncertainty in the included studies and in the assumptions made in the calculations.*

Figure 4. Summary measures of benefits and harms for an expected cohort of 10 000 women invited for 10 screening rounds from age 50 years.

## 6.2 Remaining uncertainty

Given the observational nature of this evaluation, methodological considerations have played a fundamental role when summarizing results across studies. The discussion of the individual studies in chapter 5 shows that even in the studies considered as most reliable, there are some important factors that could not be accounted for.

None of the studies could definitively address the potential influence of non-program screening on changes in breast cancer incidence and mortality, since the available information on mammography use in general and non-program screening in particular does not allow quantification of its impact. However, non-program screening

has occurred in sufficient extent to represent a potentially important source of systematic error in most studies. Estimating changes in breast cancer incidence rates in the absence of screening has been particularly challenging, partly due also to rapid changes in hormone therapy use, which is an important breast cancer risk factor. Individual level information on use of hormone therapy is not available for most of the study period, which makes accounting for this factor in the analyses more difficult.

The gradual implementation of the screening program allowed a short period with contrast in exposure (invitation for screening). Still, accounting for temporal and regional

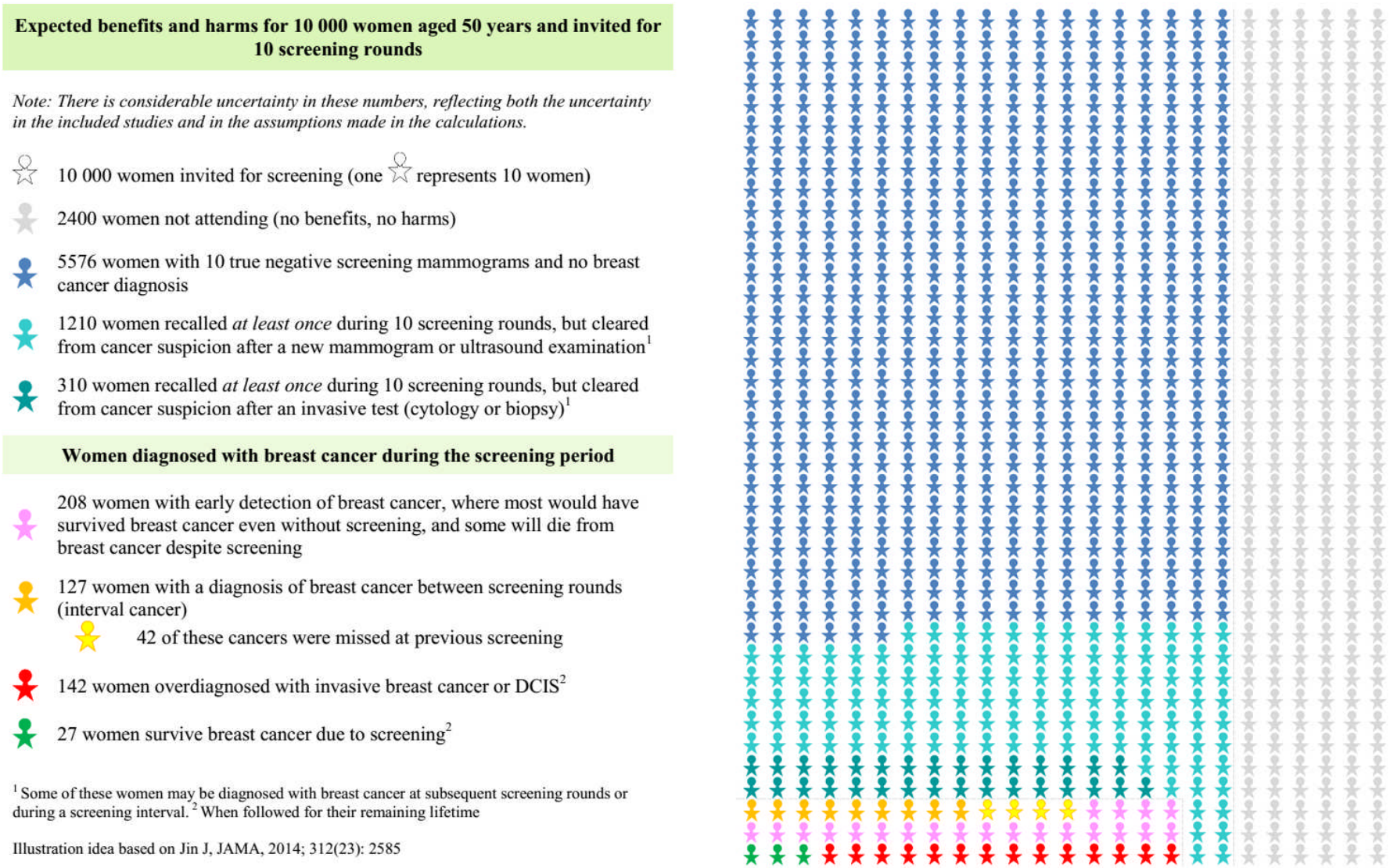



Figure 5. Summary measures of benefits and harms for and expected cohort of 10 000 women invited for 10 screening rounds from age 50 years

differences in the comparison of women invited and not invited for screening relies on unverifiable assumptions of similar trends in breast cancer incidence and mortality in the absence of screening, despite differences in breast cancer incidence and mortality prior to screening implementation.

The reorganization and centralization of breast cancer management that took place in conjunction with screening implementation could in most studies of breast cancer mortality not be separated from the effect of screening. It has been argued that this development may be responsible for a proportion of the reduction in breast cancer mortality following screening implementation [119, 169]. Effective treatment is a prerequisite for a successful screening program [53], and the benefit of each factor may therefore not easily be disentangled. Functioning breast diagnostic centers and multidisciplinary management were required before a county could be included in the screening program. As such, the NBCSP implementation may have enhanced the development towards the current organization of breast cancer management. However, in a hypothetical situation where treatment would be sufficiently effective to cure even the more advanced breast cancers, early detection through screening would be redundant. In consequence, continued improvements in treatment would be expected to reduce the absolute number of breast cancer deaths prevented through early detection provided that there is no substantial increase in incidence.

Several researchers described problems with fitting prediction models to the observed incidence rates, whereas others used crude methods with no attempts to test model fit. Uncertainty on the incidence in the absence of screening complicates the interpretation of estimates of overdiagnosis. In studies of attending and non-attending women, trend estimation would be less critical. However, these studies are vulnerable to effects of self-selection. As a consequence, none of the studies of overdiagnosis could be considered to give the most reliable estimate, and the summary estimates of overdiagnosis thus contain considerable uncertainty.

A long follow-up period is essential to quantify the full amount of benefits and harms of mammography screening. The maximum follow-up period for women invited to screening in the included studies was 14 years. In most counties, the follow-up period was much shorter. Currently, no birth cohorts have experienced all 10 screening rounds. Birth cohorts that could be observed for some years after leaving the screening program have had only a few rounds of screening. Thus, none of the studies in this evaluation were able to study the benefits and harms of screening among women who have experienced the entire screening program with subsequent life-long follow-up. Estimates of both benefits and harms may change with a longer observation period. Furthermore, there are indications that the transition from screen-film to digital mammography, which took place during most of the study period, may influence screening performance [60]. Benefits and harms estimated in this evaluation apply to a combination of these two techniques, whereas women who enter the screening program in the future will experience digital mammography only.

Finally, the estimates of cost-effectiveness depend on accurate estimation of benefits, harms and costs of screening. In the interpretation of the cost-effectiveness estimates, it should be noted that the estimated level of overdiagnosis in those analyses was lower than the summary measure provided by the steering committee. Although overdiagnosed tumors would be expected to require less extensive treatment and result in fewer quality-adjusted life-years lost than screening-detected tumors on average, higher cost-effectiveness ratios would be expected with higher levels of overdiagnosis.

In studies of the psychological effects of screening and women's attitudes towards the program, only screening attendants were included, and the results from these studies should therefore not be considered to be representative for all invited women.

## 6.3 Comparison with previous evaluations of mammography screening

The estimated effectiveness of the NBCSP in studies in this evaluation was 20-30% and indicates that the program performance is compatible with the estimates from most reviews of mammography screening trials which indicate a 20% reduction (95% CI 11 to 27%) in breast cancer mortality [14] and from the incidence-based mortality studies in the EUROSCREEN review of European population-based screening programs which indicate 26% reduction (95% CI 13 to 36%) [90].

For overdiagnosis in the NBCSP, the estimates are higher than the estimates from the Independent UK Panel on Breast Cancer Screening, which was based on the mammography screening trials. For women invited to screening and followed throughout a period after screening stops (population perspective, method B), the UK Panel considered overdiagnosis to be 5-15%, compared to 15-20% in this evaluation. The EUROSCREEN group estimated overdiagnosis compared to a situation with no screening (the population perspective, corresponding to UK Panel's method A) to be 1-10% for women followed from 50 to 79 years, which is lower than this evaluation's estimate of 15-25%.

When combining the estimates of breast cancer mortality reduction and overdiagnosis, the UK Panel concluded that there were three overdiagnoses per prevented breast cancer death [14], whereas the EUROSCREEN estimated that two deaths from breast cancer were prevented for every overdiagnosed case [170].

The Swiss Medical Board's evaluation of mammography screening in the Swiss cantons concluded that no quality-adjusted life-years were gained from mammography screening [171]. Both the methods and the input data used in the Swiss evaluation differ in several aspects from those applied in the analyses of cost-effectiveness in this evaluation. First, the evaluation of the Swiss programs was not based on data from the programs themselves. The estimated absolute effects on both mortality reduction and on overdiagnosis were smaller than the results presented in this evaluation, whereas the number of recall examinations was higher. Second, quality-adjusted life-years gained in the NBCSP were estimated for a life-time perspective, whereas the Swiss evaluation had a perspective of 13 years (6.5 years after screening ended).

The review study by Törnberg et al indicates that the rates of interval cancer in the NBCSP are comparable to those in other European population-based mammography screening programs. The probability of being recalled for further examinations are compatible with the probabilities estimated in the NHS Breast Screening Program in the UK [14] and in the EUROSCREEN review [172]. It should be noted that Norwegian data contributed to the estimates in the EUROSCREEN review. However, the recall rates in the NBCSP are much lower than estimates from the United States [173], possibly due to the younger age group and annual examinations in the United States. The reported psychological distress following false positive mammograms is consistent across studies [174, 175].

## 6.4 Questions not addressed in this evaluation

Some central aspects in the evaluation of the NBCSP have not been addressed in this report. Among these, stage shift and organizational aspects of the program were part of the evaluation's mandate, but were later excluded since none of the studies in the evaluation portfolio had investigated this.

Stage shift of breast cancer in terms of an absolute reduction in the incidence of advanced cancer (pTNM stage II and above) in a screened population compared to a situation without screening can be seen as a prerequisite for reduction in breast cancer mortality and is often used as a preliminary marker of screening effectiveness [54]. When mortality data are available, information on stage shift may be seen as less important in the evaluation of effectiveness. Studies of stage shift would also be subjected to the same challenges as studies of overdiagnosis in terms of valid estimation of the incidence rates of breast cancer in each stage in the absence of screening.

The Norwegian Radiation Protection Agency performs regular controls of the technical quality of the mammography examinations and the radiation exposure associated with mammography screening [70, 176, 177].

A major concern when balancing benefits and harms of screening is the consequences of overdiagnosis in terms of overtreatment and potential long term side effects of treatment. We are not aware of any studies in which data from the NBCSP have been used to address this. The Independent UK Panel on breast cancer screening considered excess mortality from treatment of breast cancer to be small [14]. There are, however, continuous debates on reduction in total mortality following mammography screening. Evaluation of total mortality reduction was not a part of this evaluation's assignment, and no aims of total mortality reduction were set when the NBCSP was implemented. Even though breast cancer is the most important cancer-related cause of death among women 50 years and older in Norway, breast cancer deaths constitute a small proportion of the total number of deaths in this age group. In consequence, studies of total mortality following NBCSP implementation would have limited statistical power to estimate changes precisely. In addition, the number of factors other than NBCSP invitation that could affect total mortality and thus bias the association would be much larger than for breast cancer mortality alone.

# 7 Conclusion

The committee finds that the most reliable estimate of the effectiveness of the Norwegian program indicates a mortality reduction between 20 and 30% for women aged 50-69 and followed to 79, and pertains to a situation with program screening compared to one without program screening. 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. For several reasons we consider that it is not possible to give one single point estimate for mortality reduction. The studies both outside and inside the portfolio are conducted with very different designs and analytical approaches. There was no defined plan for an evaluation of the program when nationwide implementation was decided. Thus, there were no predefined intervention and control populations where prior knowledge about important influencing factors had been taken into consideration. Such factors include regional and temporal differences in breast cancer risk and mortality, non-program screening and use of hormone therapy. If information on the expected influencing factors had been systematically collected at an individual level and regardless of screening invitation or attendance status, this would have helped the evaluation considerably. Furthermore, the implementation of the program coincided with important trends in modern multidisciplinary diagnosis and treatment of breast cancer. In consequence, based on observational studies, the evaluation could not fully distinguish between the effects of the Norwegian Breast Cancer Screening Program and the effect of multidisciplinary management.

The most important harm of the program is overdiagnosis. We consider the most reliable estimates of overdiagnosis of both invasive breast cancer and DCIS for women aged 50-79 years compared to a situation without screening (method A) to be within the range of 15% and 25%. For women aged 50-79 years in a situation with screening (method B), we consider the corresponding estimates to be within the range of 15% and 20%. The reasons for uncertainty in the estimates are the same as for mortality reduction, and additionally, the use of different denominators. The committee considers that a somewhat higher estimate of overdiagnosis in the Norwegian program, compared to estimates from reviews of the mammography screening trials, may be associated with a higher sensitivity of the Norwegian program. We emphasize that the uncertainty in these estimates is considerable.

Comparing the major benefit and major harm of screening in absolute numbers; among 10 000 invited women at age 50 to screening through 10 screening rounds, approximately 27 breast cancer deaths would be avoided at the price of 142 overdiagnosed breast cancers, or for each breast cancer death prevented, approximately 5 women are overdiagnosed. However, an overdiagnosed breast cancer would be expected to need less aggressive treatment, cause less harm and generate lower additional costs than would an additional clinically detected breast cancer.

Recall examinations after screening also comprise false positive screening mammograms. During 10 screening rounds some 20% would experience at least one recall examination due to a false positive screening mammogram. The vast majority of these women, 80%, would have only a control mammogram or an ultrasound to rule out any suspicious findings, while 20% would need an invasive test (biopsy or cytology) to exclude cancer.

False negative mammograms and interval cancers are both potential harms of the screening program. Some women will have fast growing tumors appearing in a screening interval or have tumors that are not readily detectable at mammography and these women will not benefit from the program. If help-seeking is delayed due to participation in the screening program, an interval cancer could represent harm. Every fourth breast cancer diagnosed in women attending the program is an interval cancer, of which approximately one third had a false negative mammogram from the previous screening round. These cancers were larger than true interval cancer, and the experiences of women diagnosed with interval cancer indicated that some women postponed help-seeking when they developed breast cancer symptoms, due to the negative previous mammography examination. Despite being diagnosed with interval cancer, the interviewed women remained positive towards the screening program.

The societal costs of one screening round was NOK 1389 per woman attending screening in 2012, including costs of recalls after positive mammography. The estimated 10-year treatment costs for breast cancer were NOK 356 000 measured in 2008 prices. The cost per quality-adjusted life-year gained was estimated to NOK 190 000 to 479 000, which is in the lower part of the range used by the Norwegian Directorate of Health for cost thresholds; NOK 400 000 – 1 000 000 per life-year gained.

The included studies on the women's perspective were based on the experiences of women who attended screening. They considered the invitation to screening not as an invitation, more like a call, which did not require any decision- making from them. The main motivation for participation was to get reassurance that they did not have breast cancer. Recalls were associated with mental distress, which declined over time, and the information in the recall letter was perceived as reassuring by some, and worrying by others. Women seemed to trust the screening program regardless of the recognition of false positive or false negative mammograms.

The estimated benefits and harms for 10 000 invited women aged 50 years, through 10 screening rounds, are summarized in Figure 4.

From a societal perspective, recognizing the uncertainty of the estimates, the cost-effectiveness of the program seems to be within the range of what Norwegian Health Authorities define as acceptable for health services. On the individual level, however, each invited woman has to weigh the information on potential benefits and harms based on her own values, health and life situation when deciding on whether or not to attend the program.

# 8 Recommendations for future research and evaluation

The establishment of a quality assured database that took place in parallel with this evaluation provides opportunities for future research and continued monitoring of benefits and harms associated with mammography screening. As discussed in chapter 6, there are several aspects of both benefits and harms that either could not be fully addressed in this evaluation, or may be expected to change over time. We recommend that a plan for continued evaluation and monitoring should be established. We also recommend that the extent of and results from non-program screening should be systematically and routinely reported to make such data available to health authorities and researchers.

In future studies of the NBCSP as it is organized today, a contemporaneous non-invited control group will not exist. Predicting incidence trends in the absence of screening will be even more challenging since the predictions will have to extend even further beyond the observed values than in the studies conducted so far. In consequence, comparison of attending and non-attending women may be unavoidable. Improved knowledge on factors that influence screening attendance will be critical in such studies, as will information on use of non-program screening. To reduce the influence of self-selection, information from other national registries would provide valuable contributions.

Estimates of benefits and harms for women who have experienced the full screening program (10 screening rounds) with sufficient post-screening follow-up will be particularly valuable. Such data will not be available until 2026 (at the earliest). Studies with long-term follow-up of birth cohorts of women who have not experienced the excessive use of hormone therapy during the late 1990s and early 2000s will be important in the discussions on overdiagnosis.

Furthermore, there is a need for an improved understanding of the natural history of DCIS and to what extent detection and treatment of DCIS at screening will prevent future invasive breast cancer. An increased understanding of the heterogeneity of breast cancer could be helpful in reducing harms associated with overdiagnosis of breast cancer. Well-organized translational research using clinical information from the program together with studies of samples in biobanks present a fertile ground to understand the heterogeneity further.

Future qualitative studies should also approach non-attending women to improve our understanding of reasons for not attending. The impact of a false positive mammography on re-attendance and the risk of interval cancer and breast cancer in general should also be investigated.

# 9 Definitions

*Accrual period* in the setting of incident-based breast cancer mortality is the time period during which incident breast cancer diagnoses are included.

*Bias* refers to a distortion of the effect estimate away from the true estimate due to errors in the design or conduct of the study

*Breast cancer incidence* is the number of new breast cancers during a specified time period in a defined population. Example: The number of invasive breast cancers among Norwegian women aged 50-69 years in 2012 was 1524.

*Breast cancer incidence rate* is the number of new breast cancers in a population during a specified time period divided by the total amount of person-time at risk for of developing breast cancer in the same population during the same time period. Example: The incidence rate of breast cancer for Norwegian women aged 50-69 years in 2012 was 263/100 000 person-years (1524 breast cancer cases / 580 102 person-years).

*Breast cancer mortality rate* is the number of deaths from breast cancers in a population during a specified time period divided by the total amount of person-time in the same population during the same time period. Unless otherwise specified, the number of deaths includes all deaths occurring in the population during the specified time period, regardless of when the cancer had been detected. Example: Breast cancer mortality among Norwegian women aged 50-69 years in 2012 was 35.5/100 000 person-years. This includes all deaths in women 50-69 years due to breast cancer in 2012 in Norway, regardless of whether the cancer was detected in 2012 or earlier and whether the cancer was detected while the woman was aged 50-69 or younger.

*Breast cancer risk* is the number of new breast cancers in a population during a specified time period divided by the total number of individuals in the population at the start of the time period. Also termed *breast cancer incidence proportion*.

*Case-control studies* investigate whether a specific characteristic (exposure) is more or less frequent among groups of individual who have or do not have a specific diseases or condition (outcome).

*Cohort studies* follow groups of individuals who have or do not have specific characteristics (termed exposure) and investigate whether one group is more or less likely to develop a specific disease or condition (termed outcome) when followed over time.

*Confounding* refers to a distortion of the effect estimate due to mixing of extraneous effects and the effect under study

*Dwelling time* is the amount of time that a tumor spends in each disease stage before advancing to the next stage.

*Ecologic studies* are studies in which two or more of the study factors are measured at the group or population level rather than at the individual level.

*Effectiveness* is the effect of implementing screening as a population-based program, i.e. the effect of inviting women for mammography screening.

*Efficacy* we refer to the effect of screening in woman attending screening. Efficacy should preferably be investigated in an ideal randomized controlled trial with very high attendance after invitation.

*Follow-up period* in the setting of incident-based breast cancer mortality is the time period during which deaths from incident breast cancers are counted.

*Incidence-based mortality* or *refined mortality* from breast cancer is a mortality rate that counts only the breast cancer deaths occurring among women who had their cancer detected after a specific time point, such as after screening invitation.

*Incident* screening refers to all screening rounds or screening examinations after the prevalence screening. Also termed subsequent screening.

*Interval cancers* are cancers that are detected during the screening interval in women who attended screening and had normal mammograms or normal recall investigations, or in a time period equal to the screening interval for women who have reached the upper age limit for screening.

*Lead time* is the time between the detection of breast cancer at screening and the time that the tumor would be detected if screening had not occurred, i.e. the amount of time that the date of diagnosis is advanced by screening.

*Meta-analysis* is a summary analysis of multiple studies, including statistical analyses that combine the results across studies.

*Non-organized or opportunistic screening* is examination of apparently healthy individuals at the individual's own or his/her doctor's initiative.

*Overdiagnosis* due to mammography screening is the detection of breast cancer at screening that would not have caused symptoms during the woman's lifetime, and thus would not have been detected without screening.

*Person-time* is the time at risk of developing a disease (or another event of interest) for a person followed over time. The total person-time in the denominator of a rate should be the sum of the person-time for each person in the population during the specified time period. Example: 1 person followed for 50 years and 2 persons followed for 25 years each both result in 50 person-years.

*Prevalence screening* is the first screening examination, either at a population or an individual level.

*Population attributable proportion* is the proportion of disease in the population that can be attributed to the studied factor, and depends on both the excess risk among individuals exposed to the factor and the distribution of the factor in the population. The sum of attributable proportions for several different factors can exceed 100% since most, if not all, cases of disease have more than one cause.

*Predictive value of a positive test* is the probability that an individual with a positive test truly has the disease that is being tested for.

*Randomized controlled trials (RCTs)* are experimental studies where two or more groups receive different interventions (for example treatments or health services) after random allocation of the study participants to each study group. RCTs are the highest level of evidence for effect in medical research.

*Relative survival from breast cancer* is the observed proportion of breast cancer patients still alive at a specified time after diagnosis divided by the expected proportion alive after the same amount of time in a comparable group in the general population.

*Risk ratio* is a measure of comparison of disease occurrence and is calculated as the risk of a particular event in one study group divided by the risk in another group.

*Screening* is any examination that aims to detect unrecognized disease in apparently healthy individuals.

*Screening interval* is the time between two screening examinations. In the Norwegian Breast Cancer Screening Program, the screening interval is two years.

*Sensitivity of a test* is the ability of the test to correctly identify those individuals who have the disease that is being tested for. Sensitivity is calculated as the number of true positive tests divided by the number of true positive and false negative tests and expressed as a proportion or a percentage.

*Sojourn time* is time from a breast cancer is detectable by the screening test to the time when the cancer would be detected in the absence of screening. Sojourn time is the maximum lead time.

S*pecificity of a test* is the ability of the test to correctly identify those individuals who *do not* have the disease that is being tested for. Specificity is calculated as the number of true negative tests divided by the number of true negative and false positive tests and expressed as a proportion or a percentage.

*Systematic error* refers to any distortion of the results away from the true estimate apart from random variation. Typical sources of systematic error include confounding, selection bias and information bias.

*Trend studies* investigate changes in the occurrence of a specific disease or condition over time, using aggregated data.

*Quality-adjusted life-years* is a measure of disease burden, including both the quality and the quantity of life lived.

# 10 References


1. Cancer Registry of Norway. **Cancer in Norway 2012 - Cancer incidence, mortality, survival and prevalence in Norway.** Oslo: Cancer Registry of Norway, 2014.
2. Engholm G, Ferlay J, Christensen N, Kejs AMT, Johannesen TB, Khan S, Milter MC, Ólafsdóttir E, Petersen T, Pukkala E, Stenz F, Storm HH. NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 7.0 (17.12.2014). Association of the Nordic Cancer Registries. Danish Cancer Society. Available from (09.01.2015): http://www.ancr.nu.
3. Sobin LH, Gospodarwicz MK, Wittekind C: **TNM classification of malignant tumours**. Chichester, West Sussex, UK: Wiley-Blackwell; 2009.
4. Weigelt B, Geyer FC, Reis-Filho JS: **Histological types of breast cancer: how special are they?** *Mol Oncol* 2010, **4**(3):192-208.
5. Lakhani SR, Ellis IO, Schnitt SJ, Tan PH, van de Vijver MJ (eds.): **WHO Classification of Tumours of the Breast**, 4th edn. Lyon, France: International Agency for Research on Cancer; 2012.
6. Elston CW, Ellis IO: **Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up**. *Histopathology* 1991, **19**(5):403-410.
7. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: **Molecular portraits of human breast tumours**. *Nature* 2000, **406**(6797):747-752.
8. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al*: **The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups**. *Nature* 2012, **486**(7403):346-352.
9. Sonnenblick A, Fumagalli D, Sotiriou C, Piccart M: **Is the differentiation into molecular subtypes of breast cancer important for staging, local and systemic therapy, and follow up?** *Cancer Treat Rev* 2014, **40**(9):1089-1095.
10. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, Senn HJ: **Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013**. *Ann Oncol* 2013, **24**(9):2206-2223.
11. Wist E, Naume B, Lønning PE, Schlichting E, Aas T: **Nasjonalt handlingsprogram med retningslinjer for diagnostikk, behandling og oppfølging av pasienter med brystkreft**. Norwegian Directorate of Health; 2014. Available from (01.05.2015): http://www.helsebiblioteket.no/retningslinjer/brystkreft/forord;jsessionid=A78D9536962E2E518DB528DDEC664914?hideme=true
12. Bombonati A, Sgroi DC: **The molecular pathology of breast cancer progression**. *J Pathol* 2011, **223**(2):307-317.
13. 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.

14. 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):2205-2240.
15. Cowell CF, Weigelt B, Sakr RA, Ng CK, Hicks J, King TA, Reis-Filho JS: **Progression from ductal carcinoma in situ to invasive breast cancer: revisited**. *Mol Oncol* 2013, **7**(5):859-869.
16. Silverstein MJ, Poller DN, Waisman JR, Colburn WJ, Barth A, Gierson ED, Lewinsky B, Gamagami P, Slamon DJ: **Prognostic classification of breast ductal carcinoma-in-situ**. *Lancet* 1995, **345**(8958):1154-1157.
17. Welch HG, Black WC: **Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find?** *Ann Intern Med* 1997, **127**(11):1023-1028.
18. Benson JR, Wishart GC: **Predictors of recurrence for ductal carcinoma in situ after breast-conserving surgery**. *Lancet Oncol* 2013, **14**(9):e348-357.
19. 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):2481-2484.
20. 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):1778-1784.
21. Falk RS, Hofvind S, Skaane P, Haldorsen T: **Second events following ductal carcinoma in situ of the breast: a register-based cohort study**. *Breast Cancer Res Treat* 2011, **129**(3):929-938.
22. Brennhovd IO, Capoferro R, Gundersen S, Hagen S, Harbitz TB, Høie J, Høst H, Næss A, Rosengren B, Sanderud A *et al* (eds.): **Primærbehandling Cancer Mammae**, 1st edn: Norsk kirurgisk forening; 1981.
23. Wist E, Naume B, Fjøsne HE, Lønning PE (eds.): **Blåboka. Brystkreft. Diagnostikk og behandling. En veiledning**, 7th edn: Norsk Bryst Cancer Gruppe; 2003.
24. Erikstein B, Fjøsne HE, Lønning PE, Søreide JA (eds.): **Brystkreft. Diagnostikk og behandling. En veiledning**, 6th edn: Norsk Bryst Cancer Gruppe; 2000.
25. Böhler P, Capoferro R, Gundersen S, Harbitz TB, Høie J, Høst H, Kvinnsland S, Næss A, Rosengren B, Sager EM *et al* (eds.): **Primærbehandling Cancer Mammae**, 2nd edn: Norsk kirurgisk forening. Norsk forening for terapeutisk onkologi. Den norske kreftforening; 1988.
26. Kvinnsland S, Kåresen R, Lønning PE, Søreide JA (eds.): **Cancer mammae. Diagnostikk og behandling**, 4th edn: Norsk Bryst Cancer Gruppe; 1994.
27. Erikstein B, Fjøsne HE, Lønning PE, Søreide JA (eds.): **Brystkreft. Diagnostikk og behandling. En veiledning.**, 5th edn: Norsk Bryst Cancer Gruppe; 1998.
28. **StatBank Norway**. Statistics Norway; 2015. Available from (12.01.2015): https://www.ssb.no/en/statistikkbanken
29. Falk RS: **Why are results of organised mammography screening so difficult to interpret?** *Tidsskr Nor Laegeforen* 2014, **134**(11):1124-1126.
30. Press DJ, Pharoah P: **Risk factors for breast cancer: a reanalysis of two case-control studies from 1926 and 1931**. *Epidemiology* 2010, **21**(4):566-572.
31. Veronesi U, Boyle P, Goldhirsch A, Orecchia R, Viale G: **Breast cancer**. *Lancet* 2005, **365**(9472):1727-1741.
32. Hulley S, Furberg C, Barrett-Connor E, Cauley J, Grady D, Haskell W, Knopp R, Lowery M, Satterfield S, Schrott H *et al*: **Noncardiovascular disease outcomes during**

**6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study follow-up (HERS II)**. *JAMA* 2002, **288**(1):58-66.
33. Chlebowski RT, Kuller LH, Prentice RL, Stefanick ML, Manson JE, Gass M, Aragaki AK, Ockene JK, Lane DS, Sarto GE *et al*: **Breast cancer after use of estrogen plus progestin in postmenopausal women**. *N Engl J Med* 2009, **360**(6):573-587.
34. Hofvind S, Sakshaug S, Ursin G, Graff-Iversen S: **Breast cancer incidence trends in Norway--explained by hormone therapy or mammographic screening?** *Int J Cancer* 2012, **130**(12):2930-2938.
35. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC *et al*: **Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial**. *JAMA* 2002, **288**(3):321-333.
36. Waaseth M, Bakken K, Lund E: **Patterns of hormone therapy use in the Norwegian Women and Cancer study (NOWAC) 1996-2005**. *Maturitas* 2009, **63**(3):220-226.
37. 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):130-134.
38. Hofvind S, Moller B, Thoresen S, Ursin G: **Use of hormone therapy and risk of breast cancer detected at screening and between mammographic screens**. *Int J Cancer* 2006, **118**(12):3112-3117.
39. 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.
40. Stahlberg C, Pedersen AT, Lynge E, Andersen ZJ, Keiding N, Hundrup YA, Obel EB, Ottesen B: **Increased risk of breast cancer following different regimens of hormone replacement therapy frequently used in Europe**. *Int J Cancer* 2004, **109**(5):721-727.
41. Chlebowski RT, Anderson GL: **Changing concepts: Menopausal hormone therapy and breast cancer**. *J Natl Cancer Inst* 2012, **104**(7):517-527.
42. Vennerød AM (ed.): **Norsk legemiddelhåndbok 1990-91 for helsepersonell**, 4th edn. Oslo, Norway: Norsk legemiddelhåndbok I/S; 1990.
43. **[Norwegian guidelines on estrogen treatment]**. *Nord Med* 1977, **92**(5):152.
44. Hunskår S (ed.): **Allmennmedisin - Klinisk arbeid**. Oslo, Norway: Ad Notam Gyldendal A/S; 1997.
45. Norsk helseinformatikk AS: **Norsk Elektronisk Legehåndbok.** Available from (23.03.2015): http://legehandboka.no/
46. Albrektsen G, Heuch I, Hansen S, Kvale G: **Breast cancer risk by age at birth, time since birth and time intervals between births: exploring interaction effects**. *Br J Cancer* 2005, **92**(1):167-175.
47. Keum N, Greenwood DC, Lee DH, Kim R, Aune D, Ju W, Hu FB, Giovannucci EL: **Adult weight gain and adiposity-related cancers: a dose-response meta-analysis of prospective observational studies**. *J Natl Cancer Inst* 2015, **107**(2).
48. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., Coates RJ, Liff JM, Talamini R, Chantarakul N *et al*: **Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including**

**58,515 women with breast cancer and 95,067 women without the disease**. *Br J Cancer* 2002, **87**(11):1234-1245.
49. Tretli S, Haldorsen T: **A cohort analysis of breast cancer, uterine corpus cancer, and childbearing pattern in Norwegian women**. *J Epidemiol Community Health* 1990, **44**(3):215-219.
50. Kristiansen JE: **Dette er Norge**. Statistics Norway; 2003. Available from (23.03.2015): http://www.ssb.no/befolkning/artikler-og-publikasjoner/dette-er-norge-2003
51. Jensen A: **Overvekt og fedme. Er nordmenn så overvektige?** Statistics Norway; 2007. Available from (12.04.2015): [http://www.ssb.no/helse/artikler-og-publikasjoner/er-nordmenn-saa-overvektige]
52. Præstiin LVM: **Alkoholstatistikk i 150 år. Store svingninger - klar trend.** Statistics Norway; 2008. Available from (12.04.2015): http://www.ssb.no/inntekt-og-forbruk/artikler-og-publikasjoner/store-svingninger-klar-trend
53. Wilson JMG, Jungner G: **Principles and practice of screening for disease**. Geneva: World Health Organization; 1968.
54. Vainio H, Bianchini F: **IARC Handbooks of Cancer Prevention. Volume 7 Breast Cancer Screening**. Lyon, France: International Agency for Research on Cancer; 2002.
55. Perry N, Broeders M, de Wolf C, Törnberg S, Holland R, von Karsa L (eds.): **European guidelines for quality assuance in breast cancer screening and diagnosis**, 4th edn: Office for Official Publications of the European Communities; 2006.
56. Gotzsche PC, Jorgensen KJ: **Screening for breast cancer with mammography**. *The Cochrane database of systematic reviews* 2013, **6**:Cd001877.
57. Norges offentlige utredninger: **Mammografiscreening i Norge. Masseundersøkelse for brystkreft**. **NOU 1987: 7.** Ministry of Social Affairs; 1987.
58. Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, Ljungquist U, Lundstrom B, Manson JC, Eklund G *et al*: **Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare**. *Lancet* 1985, **1**(8433):829-832.
59. **Innstilling fra sosialkomiteen om endringer i statsbudsjettet medregnet folketrygden for 1994 under enkelte kapitler under Sosial- og helsedepartementet. Innst. S. nr. 56 (1994-1995)**. 1994.
60. 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):447-455.
61. 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, 135:291-299.
62. Cancer Registry of Norway: **Mammografiprogrammet. Nøkkeltall.** Available from (04.03.2015): http://www.kreftregisteret.no/no/Forebyggende/Mammografiprogrammet/Nokkeltall/
63. Moller P, Maehle L, Apold J: **[Hereditary breast cancer]**. *Tidsskr Nor Laegeforen* 2005, **125**(22):3136-3138.
64. Moller P, Hagen AI, Apold J, Maehle L, Clark N, Fiane B, Lovslett K, Hovig E, Vabo A: **Genetic epidemiology of BRCA mutations--family history detects less than 50% of the mutation carriers**. *Eur J Cancer* 2007, **43**(11):1713-1717.

65. Moller P, Maehle L, Vabo A, Clark N, Sun P, Narod SA: **Age-specific incidence rates for breast cancer in carriers of BRCA1 mutations from Norway**. *Clin Genet* 2013, **83**(1):88-91.
66. Heimdal K, Maehle L, Apold J, Pedersen JC, Moller P: **The Norwegian founder mutations in BRCA1: high penetrance confirmed in an incident cancer series and differences observed in the risk of ovarian cancer**. *Eur J Cancer* 2003, **39**(15):2205-2213.
67. Hagen AI, Maehle L, Veda N, Vetti HH, Stormorken A, Ludvigsen T, Guntvedt B, Isern AE, Schlichting E, Kleppe G *et al*: **Risk reducing mastectomy, breast reconstruction and patient satisfaction in Norwegian BRCA1/2 mutation carriers**. *Breast* 2014, **23**(1):38-43.
68. Andersen TI, Borresen AL, Moller P: **A common BRCA1 mutation in Norwegian breast and ovarian cancer families?** *Am J Hum Genet* 1996, **59**(2):486-487.
69. Ertzaas AKO (ed.): **Mammografiprogrammet. Kvalitetsmanual**. Oslo, Norway: Cancer Registry of Norway; 2003.
70. Saxebøl G: **Røntgenunderseøkelser i Norge 1983. Absolutt, relativ og geografisk fordeling. SIS Rapport 1985:4**. Østerås, Norway: National Institute of Radiation Hygiene; 1985.
71. Saxebøl G, Lundgren LE: **Diagnostic medical x-ray examinations in Norway 1988. SIS Rapport 1990:6**. Østerås, Norway: National Institute of Radiation Hygiene; 1990.
72. Widmark A, Olsen JB: **Mammografivirksomhet i Norge. Tekniske resultater. StrålevernRapport 1995:5.** Østerås, Norway: Norwegian Radiation Protection Agency.
73. Børretzen I, Lysdahl KB, Olerud HM: **Radiologi i Noreg. Undersøkingsfrekvens per 2002, tidstrendar, geografisk variasjon og befolkningsdose. StrålevernRapport 2006:6**. Østerås, Norway: Norwegian Radiation Protection Agency.
74. Velken EB, Sager EM: **[The use of mammography in a private institute, who are the users and why do they come?]**. *Tidsskr Nor Laegeforen* 1994, **114**(13):1539-1540.
75. Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S: **Estimating mean sojourn time and screening sensitivity using questionnaire data on time since previous screening**. *J Med Screen* 2008, **15**(2):83-90.
76. Hofvind S: **En pilotstudie av mammografiaktiviteten ved to private røntgeninstitutter i Norge. Rapport 1-2006.** Oslo, Norway: Cancer Registry of Norway; 2006.
77. Hofvind S, Sanderud A: **Bruk av mammografi i Norge i 2005 og 2008. HiO-rapport 2010 nr 14.** Oslo, Norway: Oslo University College 2010.
78. Leistad L, Øren A: **Erfaringer og bruk av mammografi blant kvinner i alderen 45-54 år**. Trondheim, Norway: Sintef Teknologi og samfunn. Forebyggende helseforskning; 2010.
79. Mørland B, Lund Håheim L, Linnestad K: **SMM-rapport Nr. 4/2002. Screening for brystkreft. En oppsummering av kunnskapsstatus per august 2002**. Oslo, Norway: Senter for medisinsk metodevurdering; 2002.
80. Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, Giordano L, Frigerio A, Tornberg S, Van Hal G *et al*: **False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes**. *J Med Screen* 2012, **19 Suppl 1**:57-66.
81. Schag CC, Heinrich RL, Ganz PA: **Karnofsky performance status revisited: reliability, validity, and guidelines**. *J Clin Oncol* 1984, **2**(3):187-193.

82. Falk RS, Hofvind S, Skaane P, Haldorsen T: **Overdiagnosis among women attending a population-based mammography screening program**. *Int J Cancer* 2013, **133**(3):705-712.
83. Hagerup-Jenssen T: **Mammografiprogrammet i Norge. Evaluering av prøveprosjektet 1996-2000**. In: *Forskningsrapport.* vol. 2. Oslo: Cancer Registry of Norway; 2000.
84. Braaten T, Weiderpass E, Kumle M, Lund E: **Explaining the socioeconomic variation in cancer risk in the Norwegian Women and Cancer Study**. *Cancer Epidemiol Biomarkers Prev* 2005, **14**(11 Pt 1):2591-2597.
85. Elstad JI, Torstensrud R, Lyngstad TH, Kravdal O: **Trends in educational inequalities in mortality, seven types of cancers, Norway 1971-2002**. *Eur J Public Health* 2012, **22**(6):771-776.
86. Albrektsen G, Heuch I, Thoresen S, Kvale G: **Family history of breast cancer and short-term effects of childbirths on breast cancer risk**. *Int J Cancer* 2006, **119**(6):1468-1474.
87. Vikum E, Bjorngaard JH, Westin S, Krokstad S: **Socio-economic inequalities in Norwegian health care utilization over 3 decades: the HUNT Study**. *Eur J Public Health* 2013, **23**(6):1003-1010.
88. Hansen AH, Halvorsen PA, Ringberg U, Forde OH: **Socio-economic inequalities in health care utilisation in Norway: a population based cross-sectional survey**. *BMC health services research* 2012, **12**:336.
89. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, Lynge E, Euroscreen Working G: **Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies**. *J Med Screen* 2012, **19 Suppl 1**:33-41.
90. Etzioni R, Gulati R, Mallinger L, Mandelblatt J: **Influence of study features and methods on overdiagnosis estimates in breast and prostate cancer screening**. *Ann Intern Med* 2013, **158**(11):831-838.
91. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L: **Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review**. *Lancet Oncol* 2007, **8**(12):1129-1138.
92. Beckmann KR, Roder DM, Hiller JE, Farshid G, Lynch JW: **Do breast cancer risk factors differ among those who do and do not undertake mammography screening?** *J Med Screen* 2013, **20**(4):208-219.
93. Martin-Lopez R, Jimenez-Garcia R, Lopez-de-Andres A, Hernandez-Barrera V, Jimenez-Trujillo I, Gil-de-Miguel A, Carrasco-Garrido P: **Inequalities in uptake of breast cancer screening in Spain: analysis of a cross-sectional national survey**. *Public Health* 2013, **127**(9):822-827.
94. Lagerlund M, Sontrop JM, Zackrisson S: **Psychosocial factors and attendance at a population-based mammography screening program in a cohort of Swedish women**. *BMC women's health* 2014, **14**(1):33.
95. Brodersen J, Thorsen H: **Consequences of Screening in Breast Cancer (COS-BC): development of a questionnaire**. *Scand J Prim Health Care* 2008, **26**(4):251-256.
96. Norges offentlige utredninger: **Samfunnsøkonomiske analyser. NOU 2012: 16**. Ministry of Finance; 2012.
97. Morgenstern H: **Ecologic Studies**. In: *Modern Epidemiology.* 3rd edn. Edited by Rothman K, Greenland S, Lash T. Philadelphia, USA: Lippincott Willams & Wilkins; 2008: 511-531.

98. Rothman K, Greenland S, Lash T: **Validity in epidemiological studies**. In: *Modern epidemiology*. 3rd edn. Edited by Rothman K, Greenland S, Lash T. Philadelphia: Lippincott Williams & Wilkins; 2008: 128-147.
99. **FOR 2001-12-21 nr 1476: Forskrift om innsamling og behandling av helseopplysninger i Dødsårsaksregisteret (Dødsårsaksregisterforskriften).** Available from (01.05.2015): http://www.lovdata.no/cgi-wift/ldles?doc=/sf/sf/sf-20011221-1476.html
100. Norwegian Institute of Public Health: **Veiledning for leger: Slik skal dødsmeldingen fylles ut.** Available from (22.03.2015): http://www.fhi.no/helseregistre/dodsaarsaksregisteret/utfylling-av-dodsmelding
101. Alfsen GC, Lyckander LG: **Does quality control of death certificates in hospitals have an impact on cause of death statistics?** *Tidsskr Nor Laegeforen* 2013, **133**(7):750-755.
102. Glattre E, Blix E: **Evaluation of the Cause-of-Death-Statistics**. In: *Rapporter fra Statistisk Sentralbyrå 80/13*. Oslo, Norway: Statistics Norway; 1980.
103. Hofvind S, Sorum R, Haldorsen T, Langmark F: **[Incidence of breast cancer before and after implementation of mammography screening]**. *Tidsskr Nor Laegeforen* 2006, **126**(22):2935-2938.
104. Olsen AH, Lynge E, Njor SH, Kumle M, Waaseth M, Braaten T, Lund E: **Breast cancer mortality in Norway after the introduction of mammography screening**. *Int J Cancer* 2013, **132**(1):208-214.
105. Medical Birth Registry of Norway: **Medisinsk fødselsregister - statistikkbank**. Norwegian Institute of Public Health; 2015. Available from (23.03.2015): http://mfr-nesstar.uib.no/mfr/
106. Couto E, Qureshi SA, Hofvind S, Hilsen M, Aase H, Skaane P, Vatten L, Ursin G: **Hormone therapy use and mammographic density in postmenopausal Norwegian women**. *Breast Cancer Res Treat* 2012, **132**(1):297-305.
107. Abrahamsson L, Humphreys K: **A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density**. *Stat Methods Med Res* 2013.
108. Kerlikowske K, Zhu W, Hubbard RA, Geller B, Dittus K, Braithwaite D, Wernli KJ, Miglioretti DL, O'Meara ES: **Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy**. *JAMA internal medicine* 2013, **173**(9):807-816.
109. Rothman K, Greenland S, Lash T: **Precision and statistics in epidemiological studies**. In: *Modern epidemiology*. 3rd edn. Edited by Rothman K, Greenland S, Lash T. Philadelphia: Lippincott Williams & Wilkins; 2008: 148-167.
110. Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE: **Long-term effects of mammography screening: updated overview of the Swedish randomised trials**. *Lancet* 2002, **359**(9310):909-919.
111. Card D, Krueger AB: **Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania**. *The American Economic Review* 1994, **88**:772-793.
112. Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB, Tange UB, Blichert-Toft M, Rank F, Mouridsen H *et al*: **A model for determining the effect of mammography service screening**. *Acta Oncol* 2005, **44**(2):120-128.
113. Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB, Tange UB, Blichert-Toft M, Rank F, Mouridsen H *et al*: **Breast cancer mortality in Copenhagen**

**after introduction of mammography screening: cohort study**. *BMJ* 2005, **330**(7485):220.
114. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer EJ: **Effect of screening and adjuvant therapy on mortality from breast cancer**. *N Engl J Med* 2005, **353**(17):1784-1792.
115. Plevritis SK, Sigal BM, Salzman P, Rosenberg J, Glynn P: **A stochastic simulation model of U.S. breast cancer mortality trends from 1975 to 2000**. *J Natl Cancer Inst Monogr* 2006(36):86-95.
116. Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, Lund E: **Mammography activity in Norway 1983 to 2008**. *Acta Oncol* 2011, **50**(7):1062-1067.
117. Duffy SW, Cuzick J, Tabar L, Vitak B, Chen THH, Yen MF, Smith RA: **Correcting for non-compliance bias in case-control studies to evaluate cancer screening programmes**. *Applied Statistics* 2002, **51**(Part 2):235-243.
118. 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):1203-1210.
119. Weedon-Fekjaer H, Romundstad PR, Vatten LJ: **Modern mammography screening and breast cancer mortality: population study**. *BMJ* 2014, **348**:g3701.
120. Van Luijt PA, Heijnsdijk EAM, de Koning HJ: **Research-based evaluation of the Norwegian mammography screening programme; effectiveness, side-effects and cost-effectiveness**. Project report to the Research Council of Norway; 2015.
121. 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):3106-3112.
122. Hery C, Ferlay J, Boniol M, Autier P: **Quantification of changes in breast cancer incidence and mortality since 1990 in 35 countries with Caucasian-majority populations**. *Ann Oncol* 2008, **19**(6):1187-1194.
123. Autier P, Boniol M, La Vecchia C, Vatten L, Gavin A, Hery C, Heanue M: **Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database**. *BMJ* 2010, **341**:c3620.
124. 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.
125. Hery C, Ferlay J, Boniol M, Autier P: **Changes in breast cancer incidence and mortality in middle-aged and elderly women in 28 countries with Caucasian majority populations**. *Ann Oncol* 2008, **19**(5):1009-1018.
126. Day NE: **Estimating the sensitivity of a screening test**. *J Epidemiol Community Health* 1985, **39**(4):364-366.
127. 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):24-29.
128. Hoff SR, Klepp O, Hofvind S: **Asymptomatic breast cancer in non-participants of the national screening programme in Norway: a confounding factor in evaluation?** *J Med Screen* 2012, **19**(4):177-183.
129. Zbuk K, Anand SS: **Declining incidence of breast cancer after decreased use of hormone-replacement therapy: magnitude and time lags in different countries**. *J Epidemiol Community Health* 2012, **66**(1):1-7.

130. Wang H, Bjurstam N, Bjorndal H, Braaten A, Eriksen L, Skaane P, Vitak B, Hofvind S, Thoresen SO: **Interval cancers in the Norwegian breast cancer screening program: frequency, characteristics and use of HRT**. *Int J Cancer* 2001, **94**(4):594-598.
131. Boer R, Warmerdam P, de Koning H, van Oortmarssen G: **Extra incidence caused by mammographic screening**. *Lancet* 1994, **343**(8903):979.
132. Wang H, Hofvind SS, Thoresen SO: **[A pilot project with mammography--results from the first screening round]**. *Tidsskr Nor Laegeforen* 2000, **120**(27):3237-3240.
133. Zahl PH, Strand BH, Maehlen J: **Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study**. *BMJ* 2004, **328**(7445):921-924.
134. Lund E, Mode N, Waaseth M, Thalabard JC: **Overdiagnosis of breast cancer in the Norwegian Breast Cancer Screening Program estimated by the Norwegian Women and Cancer cohort study**. *BMC Cancer* 2013, **13**(1):614.
135. Lund E, Nakamura A, Mode N, Kumle M, JC T: **Overdiagnosis in the Norwegian Breast Cancer Screening Program – estimation based on record linkage and questionnaire information in the Norwegian Women and Cancer study**. In: Project report to the Research Council of Norway; 2014.
136. Zahl PH, Maehlen J: **Overdiagnosis of breast cancer after 14 years of mammography screening**. *Tidsskr Nor Laegeforen* 2012, **132**(4):414-417.
137. Duffy S, Michalopolous D: **Estimates of overdiagnosis in the Norwegian Breast Cancer Screening Programme**. Project report to the Research Council of Norway; 2014.
138. 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):491-499.
139. Jorgensen KJ, Gotzsche PC: **Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends**. *BMJ* 2009, **339**:b2587.
140. Hoff SR, Klepp O, Hofvind S: **Asymptomatic breast cancer in non-participants of the national screening programme in Norway: a confounding factor in evaluation?** *J Med Screen* 2012, **19**(4):177-183.
141. Solbjor M, Skolbekken JA, Saetnan AR, Hagen AI, Forsmo S: **Mammography screening and trust: the case of interval breast cancer**. *Soc Sci Med* 2012, **75**(10):1746-1752.
142. Solbjor M, Skolbekken JA, Saetnan AR, Hagen AI, Forsmo S: **Could screening participation bias symptom interpretation? An interview study on women's interpretations of and responses to cancer symptoms between mammography screening rounds**. *BMJ open* 2012, **2**(6).
143. Hofvind S, Yankaskas BC, Bulliard JL, Klabunde CN, Fracheboud J: **Comparing interval breast cancer rates in Norway and North Carolina: results and challenges**. *J Med Screen* 2009, **16**(3):131-139.
144. Hofvind S, Skaane P, Vitak B, Wang H, Thoresen S, Eriksen L, Bjorndal H, Braaten A, Bjurstam N: **Influence of review design on percentages of missed interval breast cancers: retrospective study of interval cancers in a population-based screening program**. *Radiology* 2005, **237**(2):437-443.
145. Tornberg S, Kemetli L, Ascunce N, Hofvind S, Anttila A, Seradour B, Paci E, Guldenfels C, Azavedo E, Frigerio A *et al*: **A pooled analysis of interval cancer rates in six European countries**. *Eur J Cancer Prev* 2010, **19**(2):87-93.

146. Kalager M, Tamimi RM, Bretthauer M, Adami HO: **Prognosis in women with interval breast cancer: population based observational cohort study**. *BMJ* 2012, **345**:e7536.
147. Hofvind S, Geller B, Skaane P: **Mammographic features and histopathological findings of interval breast cancers**. *Acta Radiol* 2008, **49**(9):975-981.
148. Hoff SR, Samset JH, Abrahamsen AL, Vigeland E, Klepp O, Hofvind S: **Missed and true interval and screen-detected breast cancers in a population based screening program**. *Acad Radiol* 2011, **18**(4):454-460.
149. Hoff SR, Abrahamsen AL, Samset JH, Vigeland E, Klepp O, Hofvind S: **Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography-- results from a retrospective review**. *Radiology* 2012, **264**(2):378-386.
150. Hofvind S, Geller BM, Rosenberg RD, Skaane P: **Screening-detected breast cancers: discordant independent double reading in a population-based screening program**. *Radiology* 2009, **253**(3):652-660.
151. Solbjor M: **Informasjon, valg og posisjoner ved mammografiscreening**. In: *Helsesosiologi. Analyser av helse, sykdom og behandling.* Edited by Tjora A. Oslo: Gyldendal Akademisk; 2012.
152. Roman M, Skaane P, Hofvind S: **The cumulative risk of false-positive screening results across screening centers in the Norwegian Breast Cancer Screening Program**. *Eur J Radiol* 2014, **83**(9):1639-1644.
153. Yaghjyan L, Colditz GA, Rosner B, Tamimi RM: **Mammographic breast density and breast cancer risk: interactions of percent density, absolute dense, and non-dense areas with breast cancer risk factors**. *Breast Cancer Res Treat* 2015, **150**(1):181-189.
154. Roman M, Hubbard RA, Sebuodegard S, Miglioretti DL, Castells X, Hofvind S: **The cumulative risk of false-positive results in the Norwegian Breast Cancer Screening Program: Updated results**. *Cancer* 2013.
155. Haug TT, Mykletun A, Dahl AA: **The association between anxiety, depression, and somatic symptoms in a large population: the HUNT-II study**. *Psychosom Med* 2004, **66**(6):845-851.
156. Pedersen HK: **Bør forventet levealder beregnes for små befolkninger?** Økonomiske analyser 6/2011, Statistics Norway. Available from (19.09.2014): http://www.ssb.no/a/publikasjoner/pdf/oa_201106/pedersen.pdf
157. Tangen T, Mykletun A: **Depression and anxiety through the climacteric period: an epidemiological study (HUNT-II)**. *J Psychosom Obstet Gynaecol* 2008, **29**(2):125-131.
158. Bjelland I, Dahl AA, Haug TT, Neckelmann D: **The validity of the Hospital Anxiety and Depression Scale. An updated literature review**. *J Psychosom Res* 2002, **52**(2):69-77.
159. Hafslund B, Espehaug B, Nortvedt MW: **Effects of false-positive results in a breast screening program on anxiety, depression and health-related quality of life**. *Cancer Nurs* 2012, **35**(5):E26-34.
160. Schou Bredal I, Karesen R, Skaane P, Engelstad KS, Ekeberg O: **Recall mammography and psychological distress**. *Eur J Cancer* 2013, **49**(4):805-811.
161. Osterlie W, Solbjor M, Skolbekken JA, Hofvind S, Saetnan AR, Forsmo S: **Challenges of informed choice in organised screening**. *J Med Ethics* 2008, **34**(9):e5.

162. Solbjor M, Forsmo S, Skolbekken JA, Saetnan AR: **Experiences of recall after mammography screening--a qualitative study**. *Health Care Women Int* 2011, **32**(11):1009-1027.
163. Hafslund B, Espehaug B, Nortvedt MW: **Health-related quality of life, anxiety and depression related to mammography screening in Norway**. *J Clin Nurs* 2012, **21**(21-22):3223-3234.
164. Solbjor M: **"You have to have trust in those pictures." A perspective on women's experiences of mammography screening.** In: *Researching trust and health.* Edited by Brownlie J, Greene A, Howson A. London: Routledge; 2008.
165. Solbjør M, Skolbekken JA, Østerlie W, S F: **Women's experience with mammography screening through 6 years of participation – A longitudinal qualitative study**. *Health Care Women Int* 2015, 36(5):558-76.
166. Hofvind S: **Breast cancer screening--prevalence of disease in women who only respond after an invitation reminder**. *J Med Screen* 2007, **14**(1):21-22.
167. 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):538-544.
168. 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.
169. Hofvind S, Skaane P, Elmore JG, Sebuodegard S, Hoff SR, Lee CI: **Mammographic performance in a population-based screening program: before, during, and after the transition from screen-film to full-field digital mammography**. *Radiology* 2014, **272**(1):52-62.
170. Paci E, Group EW: **Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet**. *J Med Screen* 2012, **19 Suppl 1**:5-13.
171. Swiss Medical Board, Organe scientifique: **Dépistage systématique par mammographie**. Zollikon, Switzerland; 2013. Available from (15.02.2015): http://www.medical-board.ch/
172. Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, Giordano L, Frigerio A, Tornberg S, Project E *et al*: **False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes**. *J Med Screen* 2012, **19 Suppl 1**:57-66.
173. Pace LE, Keating NL: **A systematic assessment of benefits and risks to guide breast cancer screening decisions**. *JAMA* 2014, **311**(13):1327-1335.
174. Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, Hyde C: **Systematic review of the psychological consequences of false-positive screening mammograms**. *Health Technol Assess* 2013, **17**(13):1-170.
175. Hafslund B, Nortvedt MW: **Mammography screening from the perspective of quality of life: a review of the literature**. *Scand J Caring Sci* 2009, **23**(3):539-548.
176. Hauge IH, Pedersen K, Sanderud A, Hofvind S, Olerud HM: **Patient doses from screen-film and full-field digital mammography in a population-based screening programme**. *Radiation protection dosimetry* 2012, **148**(1):65-73.
177. Pedersen K, Bredholdt K, Landmark ID, Istad TSJ, Almén A, Hauge IHR: **Teknisk kvalitetskontroll - statuskontroller for digitale mammografisystemer**. **StrålevernRapport 2010:8**. Østerås, Norway: Statens strålevern.

# 11 Appendices

# Appendix I: Objectives and framework for a research-based evaluation of the National Mammography Screening Program

## 1. Introduction

The National Mammography Screening Program is a publicly offered mammography screening of women aged 50-69 years. All women in this age group are called in for mammography screening every other year. The purpose is to discover tumours at an early enough stage to be able to provide effective treatment and thereby reduce mortality. The National Mammography Screening Program was established as a pilot project in four counties in 1995-96, becoming nationwide in 2004 as part of the National Cancer Plan (1998-2003). The Norwegian Directorate of Health and Social Services has been assigned overall responsibility for the Screening Program, while the responsibility for administration of the Screening Program has been assigned to the Cancer Registry of Norway, which is in charge of planning, carrying out, quality-assuring, and evaluating the activities. Other parties involved in the Screening Program are the National Population Register, the Norwegian Radiation Protection Authority, the Norwegian Institute of Public Health, and the regional health authorities.

The value of mammography screening has been the subject of debate, above all with regard to the measure's benefit in terms of estimated reduced mortality due to breast cancer versus its potential negative consequences and use of resources. In response to a systematic overview by the Nordic Cochrane Centre in Copenhagen (Olsen and Gøtzsche, Cochrane Database of Systematic Reviews 2002), several countries initiated evaluations of the available documentation regarding the value of screening.

In Norway, the Norwegian Directorate of Health and Social Services gave the then Norwegian Centre for Health Technology Assessment (SMM) the task of evaluating the benefit of mammography screening based on available international experience. Two Norwegian state-of-the-art reviews have been put together to summarise international screening for breast cancer. One was published as SMM Report no. 4 (2002), the other as Report no. 9-2007 from the Norwegian Knowledge Centre for the Health Services.

The first review, from 2002, was based on systematic overviews and meta-analyses of randomised studies published since 1995 and of findings from sources such as the Norwegian county mammography programs begun in 1994. For the 50-69 age group, the evaluation concluded that screening for breast cancer reduces the relative risk of dying from breast cancer by 6-27 per cent. The evidence did not indicate that overall mortality was affected. For the 40-49 age group, this report concluded that there was a balance of probability that such screening does not reduce mortality due to breast cancer.

The report from 2007 on mammography screening for women aged 40-49 is based on calculations from three systematic overviews that summarised the available effect studies and from one trial published at the end of 2006. The relative risk reduction in mortality among women urged to take part in breast cancer screening was estimated at 16 per cent after 13 years. The absolute risk reduction was approximately one fewer death per 3,000 screened women after 13 years. On the other hand, approximately ten women per 3,000 screened will be identified as having pre-malignant breast lesions (breast cancer *in situ*) which would not have developed into cancer and which the authors judged as an incorrect diagnosis of cancer followed by unnecessary treatment.

Internationally it has been suggested that the 45-49 age group may be most appropriate for inclusion in screening programs, but no review/meta-analysis of international data similar to that carried out on the 50-69 and 40-49 groups has been conducted for this age group. A review of the Swedish screening data, however, has been conducted (Nyström et al., Lancet 2002; 359: 909-19).

## 2. Overall principles of national health care policy

- A publicly administered health care system
- The entire population, regardless of age, gender, ethnic and social background, financial situation and geographical location, is to have equal access to quality health care services
- Health care services are to be medically responsible, provide high-quality care, and be adapted to the needs of users
- Efficient utilisation of resources
- Health care services are to be provided with respect for the individual's integrity and dignity

## 3. Purpose of the evaluation

The Ministry of Health and Care Services seeks an independent scientific evaluation of the National Mammography Screening Program focused on its effect on mortality due to breast cancer. One of the primary targets set for the Screening Program is to achieve a 30 per cent reduction in breast cancer-related mortality among women asked to take part in the screening. Report no. 1 (2006-2007) to the Storting states that "there is a need to evaluate the extent to which the National Mammography Screening Program has fulfilled its intentions and purpose, and to establish a scientific basis for potential expansion of the Screening Program to include other age groups."

This document outlines the objectives and framework for the research-based evaluation and establishes the underlying guiding principles governing the evaluation activities to be conducted during the evaluation period. This document was prepared by the steering group appointed to evaluate the National Mammography Screening Program, on the basis of the request from the Ministry of Health and Care Services in its letter of 22 November 2006, as well as the Research Council's dialogue with the Ministry.

## 4. Research-based evaluation

The evaluation is to be based on research, meaning it is to utilise research methodology and expertise. A call for proposals for evaluation projects will be issued on the Research Council website, and applications for project funding will be processed according to Research Council procedures. A steering group will be responsible for selecting evaluation projects for funding and for overseeing their follow-up by the Research Council. The steering group has been appointed by the Research Board of the Division for Science, and its members represent a wide range of expertise.

The evaluation projects must be largely based on existing data sources such as data found in the Cancer Registry of Norway, the Cause of Death Register, Statistics Norway (SSB) and the Norwegian Patient Register (NPR). It may be beneficial to request specially prepared data from NPR. It will be necessary to collect original data in connection with some of the evaluation projects.

Collaboration on several of the evaluation questions is encouraged, as this will facilitate better utilisation of the available data sources. Cooperation with relevant international research groups is recommended.

## 5. Topics to be evaluated

The evaluation will consist of three main topics:

- Evaluation of effectiveness of the Screening Program on mortality due to breast cancer, changes in staging, and changes in the incidence of advanced cancer
- Evaluation of the organization, availability and quality of the Screening Program as well as associated scientific development
- Economic evaluation: analysis of the combined use of resources and the benefit/effectiveness of the Screening Program

Below is a list of sub-topics exemplifying issues in need of further elucidation. This list is not meant to be exhaustive. Projects that evaluate multiple effects of the program are of special interest, as are projects assessing program organization and costs in order to shed light on the impacts of the screening.

Several of the evaluation elements below should be viewed in the context of international experience, but no new systematic reviews based upon international RCTs will be prepared under this research-based evaluation.

### 5.1 Evaluation of effectiveness

It is of interest to study time trends such as the course of mortality, incidence, and change in staging between different geographical areas and age groups, including the over-70 and under-50 groups. Studies will primarily be based on existing data sets such as registry data, and from available original studies. Pertinent issues/studies could include:

*Mortality*

- Both overall mortality and mortality due to breast cancer are of interest, even though the data set on the whole only has sufficient power for comparing mortality due to breast cancer on the basis of the pilot counties.

*Staging and histological grade*

- Analysis is needed of comparisons of changes in staging between screened and unscreened women, particularly in the case of advanced breast cancer. The use of tumour markers and other molecular-epidemiological techniques of characterising possible effects of changes in staging with a view to early diagnosis and malignancy rating also need to be explored.
- Studies analysing and clarifying issues related to true and false positive findings are especially important.
- The reliability of radiological and pathological diagnostics is also a matter of interest.

*Interval cancer*

- Studies on change in stage and type are encouraged.

*Over-diagnosis*

- The extent to which over-diagnosis occurs is greatly disputed and will be an important part of the evaluation. These issues can be studied empirically and theoretically. Epidemiological, clinical and ethical ramifications associated with over-diagnosis are also part of the challenge.

*Patient experience*

- It is important to examine and discuss patients' perceptions of availability and their experience of follow-up on diagnosed breast cancer, false positive mammographies and interval cancer.

### 5.2 Evaluation of the organization, availability, quality, and associated scientific development of the Screening Program

*Organisation, role assignments, and availability*

- Variations in attendance with respect to geographical location, age, and social indicators need to be studied, and the potential consequences of participation/non-participation need to be examined.
- Studies of why those who have been called in do not appear for screenings and patients' views and opinions on the information in the letter urging them to take part in the screening are of great interest.
- Analysis of routines for urging participation and follow-ups – as defined in the Screening Program and as they function in practice – is encouraged. Included here are also the views regarding these routines held by those working with the Screening Program as well as the target group.
- An evaluation on the roles assigned to the various players, their responsibilities and how these are upheld is also needed.

*Quality*

Topics to study include:

- The application of quality objectives described in the program guideline for quality assurance
- The use of indicators such as faulty tests, repeats, lack of re-invitations, etc.
- How quality assurance of diagnostic routines is practiced
- The scope of subsequent rounds of examinations and their outcome

*Development within disciplines and the scientific community*

Of interest is the evaluation of:

- Communication and cooperation among the disciplines, and efforts to promote scientific development
- The scope and value of evaluation and research
- Potential changes to the follow-up and management of women as a result of the Screening Program

### 5.3 Economic evaluation

An evaluation of the Screening Program in a socio-economic perspective is needed. This will generate a need to map out the economic costs to society related to implementation and administration of the program. Furthermore, the program's impacts must be evaluated in the form of societal economic cost-effectiveness and/or cost-benefit analyses.

*Use of resources*

Mapping out the costs involves identifying and quantifying the direct and indirect economic costs to society. This may encompass direct costs such as investments, administration of the Screening Program, examinations and analyses, etc., in addition to the indirect costs related to e.g. lost man-hours.

*Assessment of consequences*
The Screening Program has ramifications for such issues as mortality due to breast cancer and the quality and availability of services offered to breast cancer patients; cf. 5.1 and 5.2 above. All relevant impacts of the program need to be identified, measured and assessed.

*Economic evaluation*
In analysing the costs and cost-effectiveness of the Screening Program in a socio-economic perspective, a cost analysis of the cost-effectiveness/benefit for different age groups could be of interest. Projects are expected to be related to pertinent international studies.

**5.4 Other**
Opportunistic, unorganised screening poses another set of challenges. It may be difficult to access adequate data here. Nonetheless, there is a need to generate knowledge about the scope and frequency of examinations in various age groups, the extent of interval cancer, staging, and an estimation of the extent of over-diagnosis in connection with this type of screening activity.

## Appendix II: The Breast Cancer Database - A database created for the external evaluation of the Norwegian Breast Cancer Screening Programme

http://www.forskningsradet.no/prognett-mammografi/The_breast_cancer_database/1254009280409

# Appendix III: The Research-based evaluation of the Norwegian Breast Cancer Screening Program - Reports and publications

**189494 - Ivar Sønbø Kristiansen, University of Oslo**
**Screening costs and modelling the treatment cost for breast cancer.**
Reports

- Final report to the Research Council of Norway: The economics of mammography screening

Book/article in book/report

- Moger T.A & Kristiansen I.S. Direct and indiret cost of Norwegian Breast Cancer Screening program. Dept. Health management and Health Economics, University of Oslo, Working paper 2012:3
- Moger T.A, Bjørnelv G.M.W, Aas E. Expected ten year treatment cost of breast cancer detected within and outside a public screening programme. (*Submitted European Journal of Health Economics)*

**189488 - Jan Mæhlen/Per-Henrik Zahl/Inger Nina Farstad, University of Oslo**
Reports:

- Final report to the Research Council of Norway: Overdiagnostikk av brystkreft

Articles

- Zahl P-H, Jørgensen K.J, Gøtzsche P.C. Overestimated lead-time in cancer screening has led to substantial under-estimating of overdiagnosis. *Br J Cancer (2013) 109, 2014 – 2019*
- Zahl P-H, Mæhlen J. Overdiaggnosis of breast cancer after 14 years of mammography screening *Nor Legeforen nr 4. 2012:132; 414-7*
- Zahl P-H, Jørgensen K.J, and Gøtzsche P.C. Lead-Time Models Should Not Be Used to Estimate Overdiagnosis in Cancer Screening. *J Gen Intern Med 2014, DOI: 10.1007/s11606-014-2812-2*
- Suhrke P, Zahl, P-H. Breast cancer incidence and menopausal hormone therapy in Norway from 2004-09. A register based cohort study. (*accepted Cancer Medicine*)

**189503 - Lars Vatten, Norwegian University of Science and Technology**
**Breast cancer mortality and overall mortality rates: changes in age-specific and age-adjusted breast cancer stage distribution and histology. To compare trends in age-specific incidence and mortality from breast cancer in different European countries**
Reports

- Final report to the Research Council of Norway: Evaluation of the Norwegian Mammography Screening Program

Articles

- Weedon-Fekjær H, Romundstad P.R, Vatten L.J. Modern mammography screening and breast cancer mortality: population study. *BMJ 2014;348:g3701 doi: 10.1136/bmj.g3701 (Published 17 June 2014)*

**189504 - Siri Forsmo, Norwegian University of Science and Technology**
**Mammography Screening in Norway - The Women's Perspective**

Reports

- Final report to the Research Council of Norway: Mammography screening – the women's perspective

Articles

- Solbjør, M, Skolbekken J-A, Hagen, A I, Sætnan A R, Forsmo S. Could screening participation bias symptom interpretation? An interview study on women's interpretations of and responses to cancer symptoms between mammography screening rounds. *BMJ Open. 2012;2:e001508. doi:10.1136/bmjopen_2012-001508*
- Solbjør, M, Skolbekken J-A, Hagen, A I, Sætnan A R, Forsmo S (2012) Mammography screening and trust: the case of interval breast cancer. *Social Science & Medicine, 75:1746-1752*
- Solbjør M, Skolbekken J-A, Østerlie W, Forsmo S (2014) Women's experiences with mammography screening through six years of participation – a longitudinal qualitative study. Health Care for Women International. *DOI:10.1080/07399332.2014.989438*

Book/article in book/report

- Solbjør, M (2012) Informasjon, valg og posisjoner ved mammografiscreening. Bokkapittel til antologi om helsesosiologi. I: Tjora, A (red) Helsesosiologi. Analyser av helse, sykdom og behandling. Gyldendal Akademisk, Oslo.

**189505 - Eiliv Lund, University of Tromsø**
**Evaluations of effects: breast cancer mortality, staging and histological grade, interval cancer, overdiagnosis. Other subjects to be analyzed: opportunistic screening, use of HT**

Reports

- Final report to the Research Council of Norway: Evaluation of the project "Evaluation of the Norwegian Breast Cancer Screening Program"

Articles

- Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, Lund E. Mammography activity in Norway 1983 to 2008. *Acta Oncologica 2011; 50: 1062-1067.*
- Olsen AH, Lynge E, Njor SH, Kumle M, Waaseth M, Braaten T, Lund E. Breast cancer mortality in Norway after the introduction of mammography screening. *Int J Cancer 2013; 132: 208-214.*
- Lund E, Mode N, Waaseth M,Thalabard J-C. Overdiagnosis of breast cancer in the Norwegian Breast Cancer Screening Program estimated by the Norwegian Women and Cancer cohort. *BMJ Cancer 2013 ; 13 : 614 4.*
- Lund E, Nakamura A, Mode N, Thalabard J-C. Overdiagnosis in the Norwegian Breast Cancer Screening Program – estimation based on record linkages and questionnaire information in the Norwegian Women and Cancer study. (*To be submitted*)

**189514 - Harry J. de Koning Erasmus MC Dept. of Public Health, Nederland**
**Cost-effectiveness of breast cancer screening in Norway**.

Reports

- Final report to the Research Council of Norway: Research-based evaluation of the Norwegian mammography screennng programme; effectiveness, dise-effects and cost-effectiveness, van Luijt P.A, Heijnsdijk E.A.M, de Koning H.J.

Articles

- The research report will be adapted into two separate international publications.

**189520 – Stephen Duffy, Barts and the London School of Medicine, Wolfson Institute of Preventive Medicine, England**
**Overdiagnosis in the Mammography Breast Cancer Screening Programme in Norway**

Reports

- Final report to the Research Council of Norway: Estimates of overdiagnosis in the Norwegian Breast scressning programme, Stephen W. Duffy

Articles

- Duffy S. W, Michalopoulos D, Sebuødegård S, Hofvind S. Trends In Aggregate Cancer Incidence Rates In Relation To Screening And Possible Overdiagnosis: A Word Of Caution. *Journal of Medical Screening 2014; 21:21-9*
- Michalopoulos D, Duffy S. W. Estimation of overdiagnosis using short term trends and lead time estimates uncontaminated by overdiagnosed cases: results from the norwegian breast screening programme (*submitted J Med Screen*).

## Appendix IV. Calculation of expected number of breast cancers among 100 000 women aged 50 years and followed for 10 screening rounds.

| | | Invited women | | | | Attending women | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Screening round | Age | Number of women | Rate | Number of women at risk of breast cancer | Cases | Number of women | Detection rate | Rate of interval cancer | Number of women at risk of breast cancer | Cases detected at screening | Number of interval cancers |
| 1 | 50 | 100 000 | 266 | 100 000 | 266 | 76 000 | 0.00436 | 0.55 | 76 000 | 331 | 42 |
| 1 | 51 | 100 000 | 266 | 99 735 | 265 | 76 000 | | 1.18 | 75 627 | | 90 |
| 2 | 52 | 100 000 | 266 | 99 470 | 264 | 76 000 | 0.00436 | 0.55 | 75 537 | 329 | 42 |
| 2 | 53 | 99 000 | 266 | 99 206 | 263 | 75 240 | | 1.18 | 74 407 | | 89 |
| 3 | 54 | 99 000 | 266 | 97 942 | 260 | 75 240 | 0.00436 | 0.55 | 74 318 | 324 | 41 |
| 3 | 55 | 99 000 | 281 | 97 682 | 274 | 75 240 | | 1.18 | 73 953 | | 89 |
| 4 | 56 | 98 000 | 281 | 97 408 | 274 | 74 480 | 0.00454 | 0.55 | 73 104 | 332 | 41 |
| 4 | 57 | 98 000 | 281 | 96 134 | 270 | 74 480 | | 1.18 | 72 731 | | 88 |
| 5 | 58 | 98 000 | 281 | 95 864 | 269 | 74 480 | 0.00454 | 0.55 | 72 643 | 330 | 41 |
| 5 | 59 | 97 000 | 281 | 95 594 | 269 | 73 720 | | 1.18 | 71 512 | | 87 |
| 6 | 60 | 97 000 | 324 | 94 326 | 305 | 73 720 | 0.00573 | 0.55 | 71 425 | 410 | 41 |
| 6 | 61 | 97 000 | 324 | 94 021 | 304 | 73 720 | | 1.18 | 70 975 | | 87 |
| 7 | 62 | 96 000 | 324 | 93 716 | 303 | 72 960 | 0.00573 | 0.55 | 70 128 | 402 | 40 |
| 7 | 63 | 95 000 | 324 | 92 413 | 299 | 72 200 | | 1.18 | 68 926 | | 85 |
| 8 | 64 | 95 000 | 324 | 91 114 | 295 | 72 200 | 0.00573 | 0.55 | 68 841 | 395 | 40 |
| 8 | 65 | 94 000 | 341 | 90 819 | 310 | 71 440 | | 1.18 | 67 646 | | 84 |
| 9 | 66 | 93 000 | 341 | 89 510 | 305 | 70 680 | 0.00690 | 0.55 | 66 802 | 461 | 39 |
| 9 | 67 | 93 000 | 341 | 88 204 | 301 | 70 680 | | 1.18 | 66 302 | | 83 |
| 10 | 68 | 92 000 | 341 | 87 904 | 300 | 69 920 | 0.00690 | 0.55 | 65 459 | 452 | 38 |
| 10 | 69 | 91 000 | 341 | 86 604 | 295 | 69 160 | | 1.18 | 64 209 | | 82 |
| **Sum per 100 000** | | | | | **5692** | | | | | **3765** | **1268** |
| **Sum per 10 000** | | | | | **569.2** | | | | | **376.5** | **126.8** |

**The Research Council of Norway**
Drammensveien 288
P.O. Box 564
N-1327 Lysaker

Telephone: +47 22 03 70 00
Telefax: +47 22 03 70 01
post@rcn.no
www.rcn.no

Published by:
www.forskningradet.no/mammografi

Cover design: Design et cetera AS
Photo: Shutterstock
Print: 07 Gruppen AS/The Research Council of Norway
Copies: 200

Oslo, May 2015

ISBN 978-82-12-03424-2 (print)
ISBN 978-82-12-03425-9 (pdf)

This publication may be downloaded from
www.forskningsradet.no/publikasjoner