

The Tomosynthesis study in Bergen
The To-Be follow-up trial
Synopsis - Study 1a in the main protocol

Approved by the Regional Committees for Medical and Health Research Ethics in the South East of Norway
 (official record number 2015/424) and registered at ClinicalTrials.gov (NCT03669926)

Principal investigator: Solveig Hofvind, PhD, Cancer Registry of Norway
 solveig.hofvind@kreftregisteret.no

2020-08-04

Study title	Digital breast tomosynthesis – the future screening tool for breast cancer? Interval and subsequent round breast cancer in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography screening
Study phase	We will collect information about women screened with digital breast tomosynthesis (DBT) after originally being screened with DBT (DBT after DBT) or standard digital mammography (DM) (DBT after DM) in the To-Be 1 trial in 2016 and 2017 and are diagnosed with interval or subsequent round screen-detected breast cancer during a two years follow up period (2018-2019).
Background	Prevalent DBT has shown higher cancer detection rates and lower recall rates compared to DM. Yet, there is insufficient evidence to draw any conclusions about the downstream benefits and harms after the prevalent round of DBT in a population-based screening program and there is no data on interval and subsequent round screen-detected cancer rates and histopathologic tumor characteristics from a randomized controlled trial.
Study aim	To investigate interval cancer and histopathologic tumor characteristics for women screened with DBT or DM and subsequent round screen-detected cancer and histopathologic tumor characteristics for women screened with DBT after DBT or DBT after DM.
Study setting	The breast center at Haukeland University hospital, as a part of the national screening program, BreastScreen Norway.
Study design	Follow-up from a randomized controlled trial and a single-group clinical trial.
Outcome measures	<i>Primary outcomes:</i> Interval and subsequent round screen-detected breast cancer rates <i>Secondary outcomes:</i> Histopathologic tumor characteristics of interval and subsequent round screen-detected breast cancer <i>Other outcome measures:</i> Sensitivity, specificity, recall rates, biopsy rates, positive predictive values of recalls and biopsies
Study population	We expect 90% of the women attending the screening unit at Danmarksplass to participate in To-Be-2, based on preliminary results from To-Be 1. We therefore expect to screen 32 400 women with DBT, where 13 000 have a prior DBT, and 13 000 a prior DM. About 6400 women will be prevalently screened with DBT.
Inclusion/exclusion criteria	Inclusion criteria: Women who attend BreastScreen Norway in Bergen 2018-2019 with a complete screening examination, signed informed consent, and a prior examination in the To-Be 1 trial. Exclusion criteria: Women with breast implants, prior history of breast cancer, metastases, other primary cancers, women who report breast symptoms, and those prevalently screened with DBT in To-Be 2.
Study groups	Group 1: Women originally screened with DBT in To-Be 1, and DBT in To-Be 2. Group 2: Women originally screened with DM in To-Be 1, and DBT in To-Be 2.
Procedures	Screening with two-view DBT. Independent double reading of the screening mammograms, by a pool of seven breast radiologists. All cases with a positive

	score are discussed at a consensus meeting where the decision of whether to recall the women for further assessment will be taken.
Assessments	Women recalled will undergo further assessment, such as additional imaging and needle biopsy.
Sample size calculation	Our primary outcome of interest is the difference in rate of screen-detected cancer between the two scenarios, with a significance level of 0.05. Using a two-sided chi-squared test, group sample sizes of 13 000 in both groups achieve 80% power to detect a difference of 0.3 percentage points in screen-detected cancer rates between the two groups. This assumes the rate of screen-detected cancer in-group 1 is 0.6% under the null hypothesis.
Statistical analysis	Variables will be described and tested using chi squared tests, t-tests, one way analysis of variance (ANOVA) and Z tests. The primary outcome will be analyzed with a log-binomial regression model and presented as crude risk ratios with 95% confidence intervals.
Safety considerations	In addition to adhering to the ethical approvals obtained, the study followed a rigorous quality assurance plan with monthly reporting. The results were only available for the steering committee.
Project management	Consortium: Haukeland University Hospital Cancer Registry of Norway University of Oslo The consortium appointed a Steering Committee. The project group is led by PI Solveig Hofvind
Study sponsor	Cancer Society of Norway, Radiological department, Haukeland University Hospital and the Cancer Registry of Norway.

Digital Breast Tomosynthesis – the future tool for breast cancer screening?

Current knowledge within the field

Digital breast tomosynthesis (DBT) is a new “three-dimensional” screening tool for breast cancer, claimed to be superior to standard two-dimensional (2D) digital mammography (DM) (1-8). The sensitivity of DM is about 75% (9), and is notably lower for women with mammographically dense breasts (10). Moreover, DBT is associated with a lower recall rate, and a 30-50% higher rate of screen-detected breast cancer compared to DM (1-8). However, a higher rate of screen-detected breast cancer is beneficial for women and society only if the detected tumors are small and have aggressive characteristics (“killing cancers”) as slow growing tumors might represent overdiagnosis and result in overtreatment (11, 12). If DBT detects the “killing cancers”, we expect a lower rate of interval cancer and advanced cancer in subsequent screening rounds compared to DM.

As of today, there is inadequate evidence to support the use of DBT for screening. No studies have reported interval cancer rates or prognostic tumor characteristics based on DBT in screening programs run according to the European guidelines for breast cancer diagnosis and treatment (13), i.e. organized programs with independent double reading and biennial screening for women aged 50-69. The majority of studies on DBT are retrospective, based on a single screening examination using equipment from one vendor, and are performed in the U.S. where annual screening is recommended from the age of 40 (5-8, 14). Further, most studies have evaluated DBT in combination with DM (DBT+DM), which almost doubles the radiation dose given to women (15). Other issues with inadequate evidence to support decision-making are the expected increase in examination and interpretation time; increased burden on IT systems; and the power, speed and economic costs of running a screening program, including work up and treatment, with DBT versus DM.

Implementing a new screening technique carries expected and unexpected challenges. Before any conclusion of efficacy can be drawn, the benefits must be balanced against the disadvantages, particularly in a population-based program where asymptomatic women are screened. Even with the use of DBT, 2D images are essential for the perception of e.g. microcalcifications and for comparison in the next screening round. Whether the radiation dose associated with a double exposure (DBT+DM) is acceptable might be questionable. Studies have shown that synthetic mammography (SM), which is a 2D image reconstructed from raw data obtained during the DBT exposure, can replace DM (2, 16). Unlike DM, SM does not involve exposing women to additional ionizing radiation. However, more information on radiation doses and early performance measures, including recall rates and rates of breast cancer for DBT+SM, DBT+DM, and DM alone from different vendors, are needed.

The higher detection rate of screen-detected breast cancer among women examined with DBT+DM versus DM might be due to learning and/or a prevalence effect. However, studies to date present only a cross-sectional view of DBT effectiveness. This does not reflect the continuum that characterizes organized mammographic screening where screening is recommended every other year for 20 years according to European guidelines (13). In Norway and most other European countries, women are invited to screening ten times from the age of 50. In the US, women often have a yearly mammogram from the age of 40, translating to 30-40 screening examinations during a lifetime. It is therefore essential to investigate the effects of DBT for more than a single examination.

The substantial increase of screen-detected breast cancers among those screened with DBT+DM versus DM makes us assume that there will be a corresponding decrease in the interval cancer rate and in the rate of advanced screen-detected breast cancer in later screening rounds. At least two years of follow-up is required to estimate interval cancer rates in biennial screening programs. Due to limited follow up in the studies performed, only one study with an annual screening has reported

interval cancer rates (17). No decrease was observed, however, this might be due to small numbers (underpowered study) and the frequency of screening in the study setting.

Another aspect of DM and DBT is compression and pain (18, 19). During mammography, the breast is compressed between a detector and paddle to improve image quality and reduce radiation dose. Mammography vendors suggest optimal values for compression force, but these are not evidence-based (13) or adapted for DBT. A systematic review has shown that 25-46% of women may not attend their next screening appointment because of breast compression-related pain (20). As far as we are aware, only two studies have reported results on the association between breast compression and mammographic sensitivity (21, 22). There is a considerable need to address this lack of knowledge.

The Tomosynthesis trial in Bergen (TOBE trial) is a randomized controlled trial investigating whether DBT+SM is superior for breast cancer screening than DM (<https://clinicaltrials.gov> and <https://www.krefregisteret.no/screening/Mammografiprogrammet/TOBE-studien>). The study started in October 2015 and is currently ongoing. All women attending the Norwegian Breast Cancer Screening Program at Danmarks plass in Bergen during 2016-2017 are invited to participate. The TOBE-trial is the first and largest study evaluating DBT+SM using equipment from another vendor (GE) as the majority.

TOBE is an ongoing randomized trial investigating the efficacy of DBT+SM versus DM in an organized population based screening program run according to the European guidelines. Approximately 35 000 women aged 50-69 will be asked to participate in the trial and preliminary results indicate a 90% participation rate. Women participating in the trial are randomized using a 1:1 allocation ratio to screening with DBT+SM or standard DM. The trial is a collaboration between the Cancer Registry of Norway, Haukeland University Hospital, and the University of Oslo. The study is financed by the Research Council of Norway and has received funding of 12.3 million NOK over a 5-year period (2015-2019). The TOBE trial is answering research questions related to radiation doses from DBT+SM versus DM, interpretation time, IT storage requirements, recall rates, screen-detected breast cancer rates, and cost-effectiveness. Information about interval breast cancer will be available in 2020, two years after the last woman is screened. Although the rates will be small, our applied study will be the first randomized trial analyzing data on interval cancer rates for DBT+SM.

Objectives of the project

Results from the ongoing TOBE trial will fill some of the knowledge gaps described above. However, the first year of running the TOBE trial and recent publications on the topic have identified additional challenges and new evidence gaps that are important to address before DBT can be considered for use in organized screening. This view is supported by several review studies (23, 24), the IARC Handbook in Breast cancer Screening (25), and the recommendations given by the European commission in breast cancer (26).

The TOBE trial is a well-planned and well-executed randomized controlled trial, comparing selected early performance measures achieved in screening with DBT+SM with results from standard DM. The TOBE-trial consortium (steering committee) meets regularly to ensure the progress and quality of the trial. This application argues to extend the TOBE trial for a total of five years (TOBE-2). This extension will consist of one additional screening round (two years) and follow-up for three years. To investigate the effect of subsequent screening with DBT+SM we want to screen all women attending screening at Danmarks plass, Bergen with DBT+SM in 2018 and 2019. This would allow us to analyze data on interval breast cancer among women screened with DBT+SM after DBT+SM and with DBT+SM after DM in the original trial (hereafter referred to as TOBE-1) in 2020. Continuing TOBE-1 with TOBE-2 in January 2018 is the only opportunity to get information from women subsequently screened with DBT+SM, and have a prior DBT+SM or DM based on random allocation.

We will conduct a retrospective review of interval cancers from TOBE-1 and screen-detected cancers in TOBE-2 and classify these cases as “missed” or “true” based on the prior mammograms. This will let us estimate possibly over- and underdiagnosed cases and understand more about the efficacy of DBT+SM (referred to as DBT only from here on in) in a screening setting. This extension of the TOBE-1 trial is necessary to provide new knowledge about subsequent screening rounds. The established IT systems, procedures and workflow from TOBE-1 make it easy and realistic to prolong the study while minimizing costs, to gain the additional knowledge needed to fully evaluate the merits of DBT. Specifically, in TOBE-2 we will evaluate the five themes described below.

Study I, Early performance measures for screening with DBT after DBT, and DBT after DM. The rationale for Study I of TOBE-2 is to investigate the effect of subsequent screening with DBT. This study consists of two parts: a) early performance measures and b) possible learning effects. Early performance measures of interest are the recall rate and rate of screen-detected breast cancer, positive predictive values, and prognostic and predictive histopathological tumor characteristics. These measures will be compared for DBT after DBT, and DBT after DM alone. We expect a continuation of the previously observed low recall rate and a healthy (sustained or still high relative to DM alone) rate of breast cancer for women screened with DBT after DBT compared with women screened with DBT after DM alone. We will consider early performance measures beneficial for DBT after DBT versus DBT after DM if the low recall rate persists and the rate of breast cancer decreases, particularly for advanced disease, given the expected increase in detection rate for DBT versus DM in TOBE-1.

Since all radiologists participating in TOBE-2 are experienced in reading DBT from TOBE-1, we will investigate possible learning effects by comparing e.g. interpretation time for women screened in TOBE-2 with results from TOBE-1 and selected early performance measures for women screened with DBT after DM. Any possible learning curves, for example with interpretation time, are expected to dissolve in TOBE-2. Study I will result in two papers, “Subsequent screening with DBT – results from a randomized controlled trial performed in a population based screening program” and “Possible learning effects from implementation of DBT in screening for breast cancer” which will be submitted to Eur J of Cancer (Impact factor 5.5) and Eur Radiology (Impact factor 4.0), respectively.

Study II, Interval breast cancer following screening with DBT versus DM will focus on interval breast cancers identified among women screened in TOBE-1. We will analyze rates and prognostic and predictive tumor characteristics. Analyses will be stratified by mammographic density.

Interval breast cancers are a shortcoming of mammographic screening because they have less favorable prognostic and predictive histopathological tumor characteristics than screen-detected cancers (27-32). Interval cancers are a challenge for screening programs because they decrease program sensitivity and contribute substantially to breast cancer mortality in the screened population. Moreover, 3-35% of interval cancer cases represent findings that were detectable but overlooked at the time of screening. This may increase women’s distrust in mammographic screening (33-35). The expected increase in the rate of screen-detected breast cancer observed with DBT screening suggests DBT will have lower interval cancer rates than DM. We may similarly expect a lower rate of advanced interval cancers associated with DBT screening versus DM, as we expect tumors to be detected earlier at screening. However, these hypotheses have yet to be verified.

The design of this study is unique for estimating the interval cancer rate. The randomization in TOBE-1 offers the best opportunity for a valid comparison between the study and control groups. The databases at the Cancer Registry of Norway (36) ensure valid and complete data about interval cancers. A unique PIN for each individual makes it possible to identify all breast cancers among women screened, including women who have moved. This study is also unique internationally due to high data completeness, opportunities for linkage and its randomized design. We will submit an

article “Interval breast cancer in a randomized controlled trial with DBT and DM” to Radiology (Impact factor 6.9).

Study III focuses on missed and true screen-detected and interval breast cancer in mammographic screening with DBT versus DM. This study is a retrospective review of prior mammograms from women with interval and screen-detected breast cancers detected in TOBE-1 and -2, respectively. Cases will be classified as missed or true. This will let us evaluate potentially over- and underdiagnosed cases.

The purpose of this study is to identify missed and true interval and screen-detected breast cancers and to classify their mammographic features in order to learn why they were missed. No studies using DBT have been published on this topic. We expect the consequence of missing cases is fatal, but recent studies indicate that true interval cancer cases detected with DM are more aggressive than missed cases because they are more frequently of the triple-negative subtype (37). Results from our study, “Organized breast cancer screening with digital breast tomosynthesis: Missed and true interval and screen-detected breast cancer”, will be submitted to Lancet Oncology (Impact factor 24.7).

In Study IV, “Expected and experienced discomfort and pain in DBT by compression force and pressure”, we will collect information about women’s expectations and experience of discomfort and pain in mammographic screening with DBT. The rationale of this study is to explore whether individualized, standardized compression pressure influences women’s screening experiences. Compression force refers to the weight applied to the breast, while pressure takes into account both force and the breast area being compressed. We will use a compression paddle that indicates when a pressure of 10 kilopascal has been reached on one screening unit while the other will use a standard paddle with no such indication. Results of the study, “Expected and experienced discomfort and pain in mammography screening with DBT”, will be submitted to Cancer (Impact factor 5.6).

We will perform an economic evaluation of continuous use of DBT in Study V. The study will investigate the costs of DBT after DBT versus DBT after DM. The rationale of the study is to estimate the financial impact of running a screening program with DBT in an everyday setting. A major advantage of this study is the availability of individual level data on all diagnostic and treatment components, cancer stage distributions and recall rates within the project in a subsequent screening round. We will submit a paper, “Economic costs of using DBT in an everyday setting in a population based screening program” to The Eur J of Health Economics (Impact factor 2.3).

Methodological approach and calculation of statistical power

As a continuation of TOBE-1, we will invite all women attending screening at Danmarksplads after 1.1.2018 to participate in TOBE-2. Women who consent will be screened with DBT according to standard procedures. Those who decline will be screened with DM. Seven breast radiologists will be involved in reading images for this study. As with TOBE-1, we will not run any interim analyses in order to avoid influencing the staff who work with the equipment, particularly radiologists. This is a unique decision.

We expect 90% of the women attending the screening unit at Danmarksplads to participate in TOBE-2, based on preliminary results from TOBE-1. We therefore expect to screen 32 400 women with DBT, where 13 000 have a prior DBT, and 13 000 a prior DM. About 6400 women will be prevalently screened with DBT (Table 1). Information from databases at the Cancer Registry will be merged to obtain women’s individual screening history in the program. Data about screening participation, radiologists’ interpretations, screening outcomes, results of any recall examinations, and mammographic features for participants in TOBE-2 will be collected and stored according to the procedures developed for TOBE-1. Data on compression, mammographic density and radiation dose is considered part of the screening data.

Table 1: Number of women in the target group of TOBE-2, assumed number of participants, recalled women and breast cancer cases

	Total	DBT after DBT	DBT after DM	Prevalent screened
Women invited to Danmarksplads (n)	48 000	-	-	-
Expected participants at screening*	36 000	14 500	14 500	7000
Expected participants in TOBE-2**	32 400	13 000	13 000	6400
Expected recalls***	975 (3%)	310	310	155
Expected screen-detected cancers***	200 (6/1000)	80	80	40
Interval breast cancer***	62 (2/1000)	25	25	12

*75% participation rate in the program, **90% participation rate in TOBE-2, ***Based on results from 2013 and 2014

Real-time data tracking is needed to keep study results updated and the study operating according to stated timelines. Clinical data related to recall examinations will therefore be collected immediately after diagnosis. The research assistants for TOBE-1 have established high quality data handling procedures and preliminary analyses show 100% data completeness. A close collaboration with the pathologists at Haukeland University Hospital, led by Professor Akslen, has been established. The TOBE-1 research assistants will continue coding and maintaining complete pathological information for TOBE-2. Data needed to perform the outlined studies will thus be available about two weeks following the final diagnosis, after surgical treatment.

Study I, early performance measures after subsequent screening with DBT: We will assess differences in early performance measures (Study Ia), and learning effects including interpretation time (Study Ib), measured as proportions, means, medians, and ranges for women who have been screened with DBT after DBT (group 1) versus DBT after DM (group 2). We will evaluate statistically significant differences using chi-squared tests for proportions and rates, t-tests for means, Mann-Whitney tests for medians, and ANOVA, assuming a type I error rate of 0.05.

Our primary outcomes of interest for study Ia are differences in recall rates and rate of screen-detected cancer between the two screening scenarios, with a significance level of 0.05. Using a two-sided chi-squared test, group sample sizes of 13 000 in both groups achieves 80% power to detect a difference of 0.6 percentage points in recall rates between the two groups. This assumes the recall rate in group 1 is 3% under the null hypothesis (Table 1). Similarly, using a two-sided chi-squared test, group sample sizes of 13 000 in both groups achieve 80% power to detect a difference of 0.3 percentage points in screen-detected cancer rates between the two groups. This assumes the rate of screen detected cancer in group 1 is 0.6% under the null hypothesis (Table 1).

The main outcome of study Ib is change in interpretation time for DBT. The primary analysis will evaluate the mean difference in interpretation time for subsequent versus prevalent DBT screens. This study will include all women with two DBT screens (prevalent and subsequent). We will use one-way analysis of variance (ANOVA) to determine whether there was a change in the mean difference in interpretation times for any radiologist. The likelihood of a radiologist interpreting both examinations for a given woman is about 50%; in cases where the same pair of radiologists have interpreted both screens, one reader will be chosen at random to be included in the study. Approximately 13 000 women in TOBE-2 will have been screened with DBT in TOBE-1, thus $13000 \times 0.5 = 6500$ women will be included in this analysis. Assuming seven readers have each read 928 pairs of eligible images, and within-rater variance in interpretation time is 25, this study has 80% power to detect a change in mean interpretation times of less than one second at a significance level of 0.05 if between-group variances are ≥ 0.06 . Our study is well-powered for this analysis.

In *Study II* we will use interval cancer data from women screened in TOBE-1. All women screened in TOBE-1 (2016-2017) will be followed for interval breast cancer for up to 2 years during TOBE-2. Interval cancer data will therefore be available during the first quarter of 2020.

The primary outcomes of interest for this study are the rates of interval breast cancers and the proportions of interval cancers relative to all detected breast cancers for both study arms. We will additionally analyze rates and proportions of prognostic and predictive tumor characteristics of interval cancers following screening with DBT+DM versus DM only. Complete information about these characteristics will allow us to identify features that might be typical for interval cancers following screening with these two modalities. We will use information about hormonal receptor status, Her2Neu and Ki67 to classify tumors by immunohistochemical subtypes (38). Early stage Luminal A tumors might indicate tumors with low progression compared with late stage, Her2+ or TN tumors. Differences in the distributions of the characteristics between these groups may further indicate whether the expected extra tumors detected with DBT represent tumors with slow progression compared with those detected with standard DM. Descriptive analyses including frequencies, means and medians will be used to compare rates and percentages.

Interval cancer rates are low, approximately 2/1000 screens (Table 1). Overall, we expect to observe approximately 25 interval cancers in each arm of the TOBE-2 study. Our study is thus underpowered to test for differences in rates or proportions of interval cancers between study arms. With respect to tumor characteristics, our study is underpowered to observe differences in the proportion of a given tumor characteristic of less than 38% between study arms. However, the results of this study will be of substantial interest due to the lack of data on this topic and the randomized study design. This study will contribute data to the project “Effectiveness of digital breast tomosynthesis (3D-mammography) in population breast cancer screening – collaborative individual participant data (IPD) meta-analysis”, headed by professor Houssami at the University of Sydney.

In *Study III*, a review of the screen-detected and interval breast cancers, we will perform a blind and an informed review with two internal and three external breast radiologists experienced in reading DBT. The blind review will include mammograms from all cancer cases detected in TOBE-2 as well as a random selection of false positive and negative screening examinations, and will be performed individually using procedures similar to those used in a normal screening setting. We expect about 80 screen-detected and 25 interval breast cancers in each arm (Table 1). Including a similar number of false positive and negative results in a total of about 200 cases, which will take about 3h to read individually (about 1 minute each examination), 6 hours in total for each of the radiologists. The informed review will be performed only for the cancer cases, which is expected to include about 2×100 cases, requiring about two-days’ work for each member of the consensus group. It will be based on all available mammographic and histopathologic information. The reviews are aimed at classifying the cases into mammographically missed, minimal sign, and true interval and screen-detected breast cancer. Information about mammographic features will be registered during the consensus review and analyzed according to features, with the aim of retrospectively identifying the cancers. Histopathologic tumor characteristics will be stratified using the mammographic classification groupings. The blind review is expected to classify a substantially higher number of cancer cases than observed in a normal screening setting; however, this review will allow us to examine reader sensitivity. To this end, the proportion of interval cancers missed by each reader during their individual review will be tabulated based on the results of the informed review.

We will use descriptive analyses (frequencies, means, standard deviations, etc.) to report the characteristics of detected cancers. Additionally, the proportion of true and missed breast cancers from each of the blind (individual) reviews will be compared to those from the informed (consensus) review (assumed gold standard) using the Kappa statistic to investigate how well individual rates agree with the consensus assessment. This will be the first radiologic review study conducted on DBT images and is therefore a hypothesis-generating study. Because hypothesis testing will not be conducted as a part of this study, a sample size calculation is not required.

In Study IV, about discomfort and pain, we will invite women attending the screening at Danmarksplass during the fall 2018, to participate. Consenting women will be requested to fill in a questionnaire about their expectations and experience of discomfort and pain during DBT using a Numeric Rating Scale (NRS) (39) (Figure 3) before and after screening. The questionnaire will also collect information on covariates such as age, weight, height, screening history, and willingness to be screened in the future.

Women will be randomized (1:1 allocation ratio) to receive screening at Unit A or Unit B. Screening at Unit A will be performed with a compression paddle installed specifically for this study that provides information about compression force and pressure for each woman (study paddle). Screening at Unit B will be performed using a standard paddle that indicates only compression force. A unique screening ID-number written on the questionnaire will be used to link questionnaire data to information about routinely recorded technical parameters associated with screening examinations, such as compression pressure, compression force, breast size, density, and compressed breast thickness.

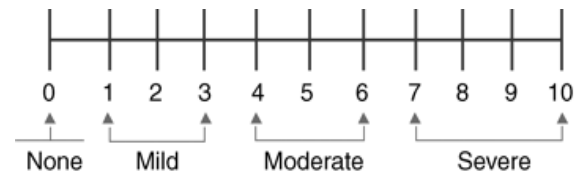


Figure 1. Commonly used one-dimensional pain intensity scales: the 11-point numeric rating scale (NRS) (30)

We expect 6500 women will attend screening at Danmarksplass during fall 2018 and 65% participation based on preliminary data from a previous questionnaire administered through the screening program. We therefore anticipate that 2080 women will be randomized into each of the study arms, for a total of 4160 participants. Descriptive analyses will be conducted to ensure the validity of the randomization. Multivariate ordinal regression (proportional odds model) will be used to determine whether women experience less pain when imaged with the study paddle compared to the standard paddle, adjusting for expected pain, screening history, use of hormonal therapy, and other acquisition and demographic characteristics of interest. The Brant test will be used to test the proportional odds assumption (40).

Assuming that women's pain scores on the NRS are normally distributed between 0-10, our study has 82.4% power to detect an odds ratio of 1.17 for a two-tailed two sample comparison of ordinal NRS scores in a univariate proportional odds ordinal model with a significance level of 0.05 (41).

Study V, economic evaluation of DBT in an everyday screening setting, will consider costs related to use of DBT after DBT versus DBT after DM in a screening setting and will use the same strategy as in TOBE-1 for data collection. This implies that estimates on the costs of DBT after DBT versus DM will be based on the investment and for the recorded time used for each patient. All interventions within the project, whether diagnostic or therapeutic, performed on the patients during follow-up are routinely collected with relevant procedure codes for radiology, medicine and surgery, as well as the name of prescribed medications. The approach avoids waiting for data on resource use from the Cancer Registry, the Patient Registry and the Prescription Registry, which would lead to a considerable time lag before data could be analyzed. We will estimate costs for each procedure using either radiology reimbursement weights (diagnostic procedures), diagnostic related groups (DRG) codes cost weights (for in-hospital treatment such as surgery, radiotherapy and chemotherapy) or retail sales prices (hormonal therapy). These analyses will be solely descriptive; a power calculation is thus not needed.

Project plan including tentative milestones

Professor Hofvind will be responsible for the TOBE-2 trial. She is the PI of TOBE-1 and has substantial knowledge about the screening program and its databases. Prof. Hofvind has extensive experience leading studies and she will supervise the Post Doc position applied for as a part of this study. Prof. Aslaksen is co-PI and heavily involved in TOBE-1 and will do his best to run TOBE-2

as well and efficiently as possible. He is also the head of the steering committee of TOBE-1, and is willing to continue this role in TOBE-2. Dr. Nerås is the medical leader of the breast centre at Haukeland and will be responsible for the practical radiologic aspects of the trial, while Prof. Akslen will oversee histopathologic activities. Associate Prof. Moger will lead the cost analyses performed in Study V. Moger was highly involved in the economic aspects of the external, research-based evaluation of the screening program. Prof. Skaane, led the Oslo DBT trial and will play an important role in the data analyses and interpretation, together with associate Profs. Zackrisson and Lee, and Prof. Houssami. Lee and Zackrisson are practicing breast radiologists and heavily involved in academic radiology. Zackrisson is the PI of a DBT trial in Malmo. Prof. Houssami is co-leading the Screening with Tomosynthesis or Standard Mammography (STORM) trial and wrote a chapter about screening modalities in the International Agency for Research on Cancer Handbook of breast cancer screening (25). The research assistants working at the Cancer Registry and the breast centre at Haukeland (Holen, MSc and Hanestad. BSc), are responsible day-in and day-out for all data collection/completion, quality assurance, reporting, and preliminary analyses for TOBE-1. They will continue to be key players in this role in TOBE-2. Additionally, Sebuødegård (MSc) and Lilleborge (PhD) are statisticians with extensive knowledge in data management and the screening database. They will have a central role in data extraction and cleaning for all studies, and will support the statistical components for all studies, including that performed by the Post Doc.

Studies I, II and III will be a part of a post doc project. Study IV will be led by the PI and Dr. Moshina and will be run in close collaboration with the radiographers and research assistants. Associate professor Moger at the University of Oslo will lead Study V, which will be performed by a researcher.

Table 2: Time schedule and milestones for the planned project

Start	End	Activity
Aug 2017	Dec 2017	Prepare TOBE-2
Jan 2018	Dec 2019	Data collection for TOBE-2
Jan 2018	Dec 2022	Data completion, quality assurance and preliminary analyses I, II, III, V
Aug 2018	Dec 2018	Data collection for Study IV
Jan 2019	Dec 2019	Data cleaning, analyses and writing Paper IV (PI)
Jan 2020	-	Hire post doc
Jan 2020	Mar 2020	Prepare data for analyses (studies I, II, III, V)
Jan 2020	Dec 2020	Study Ia: Data cleaning, analyses, writing and submission, Paper Ia (post doc)
Jan 2020	Dec 2021	Study V: Data cleaning, analyses, writing and submission, Paper V (researcher)
Aug 2020	Mar 2021	Study Ib: Data cleaning, analyses, writing and submission, Paper Ib (post doc)
Aug 2020	Dec 2021	Study III – performing the review (post doc)
Jan 2021	Dec 2021	Study II: Data cleaning, analyses, writing and submission, Paper II (post doc)
Jan 2022	Dec 2022	Study III: Data cleaning, analyses, writing and submission, Paper III (post doc)

Dissemination and communications strategy

TOBE-2 will put Bergen and Norway in the spotlight internationally due to our execution of well run randomized controlled trials. All studies will be submitted to high impact peer-reviewed international journals in the fields of radiology, breast oncology, epidemiology, and/or health economics. The Cancer Society of Norway will be acknowledged in all publications. The research group will encourage the post doc, the researcher and other participants of the project group to submit abstracts to national and international congresses in order to build a network within the field of radiology, specifically mammography, and screening with DBT. We will also publish lay summaries of the project and individual studies to share results with women invited to screening, the public, professionals involved in the screening program, health politicians, and other stakeholders. We will use the homepage of the Cancer Registry of Norway to inform women targeted by the screening program and the general population about our project. A press release will be made when individual papers are available online or in print.

Ethical considerations

TOBE-1 is REK-approved (Reference 569184) and an application to extend the trial (TOBE-2) will be submitted in June 2017. Data collection for TOBE-2 will be performed based on protocols developed for TOBE-1, according to the Cancer Registry regulations and Helseregisterloven §8, 7th edition (38). The linkage key between the PIN and ID will be stored at the Data Delivery Unit at the Cancer Registry. Only information from women who have signed an informed consent will be used. The post doc and researchers will receive de-identified data for all analyses.

Why should this study be supported?

DBT is touted as next generation screening technology for breast cancer (42). However, there is insufficient evidence about the effects of DBT to justify implementing this technology in a screening setting. TOBE-2 provides a unique opportunity to fill several knowledge gaps needed to assess the technology's potential as a tool for organized screening, both in Norway and internationally. TOBE-2 covers all parts of a health technology assessment by including medical evaluation studies (Studies I, II and III), ethical/patient choice evaluation (Part of Study I (radiation doses) and Study IV) and cost-evaluation (Study V).

Our project group includes highly qualified doctors and academics with directly relevant experience with DBT, epidemiology, statistics, and health technology assessment. This multi-disciplinary team will ensure TOBE-2 is run effectively to a successful completion.

All Norwegian women aged 50-69 are invited to participate in the screening program. The effectiveness of the program is measured by its effect on society (among invited women) and per protocol (participants).

References

1. Ciatto S et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *The Lancet Oncology*. 2013;14(7):583-9.
2. Bernardi D et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *The Lancet Oncology*. 2016;17(8):1105-13.
3. Skaane P et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology*. 2013;267(1):47-56.
4. Lang K et al. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. *European radiology*. 2016;26(1):184-90.
5. Friedewald SM et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *Jama*. 2014;311(24):2499-507..
6. Durand MA et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology*. 2015;274(1):85-92.
7. Mariscotti G et al. Digital breast tomosynthesis as an adjunct to digital mammography for detecting and characterising invasive lobular cancers: a multi-reader study. *Clinical radiology*. 2016;71(9):889-95.
8. Rafferty EA et al. Breast Cancer Screening Using Tomosynthesis and Digital Mammography in Dense and Nondense Breasts. *Jama*. 2016;315(16):1784-6.
9. Hofvind S et al. Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. *European journal of epidemiology*. 2007;22(7):447-55.
10. Boyd NF et al. Mammographic density and the risk and detection of breast cancer. *The New England journal of medicine*. 2007;356(3):227-36.
11. Puliti D et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *Journal of medical screening*. 2012;19 Suppl 1:42-56.
12. The benefits and harms of breast cancer screening: an independent review. *Lancet*. 2012;380:1778-86.
13. Perry N et al. European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembourg: Office for Official Publications of the European Communities, 2006. ISBN 92-79-01258-4: 2006. Cited 20/05/2017
14. Lee CI et al. Comparative effectiveness of combined digital mammography and tomosynthesis screening for women with dense breasts. *Radiology*. 2015;274(3):772-80.

15. Svahn TM et al. Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. *Breast (Edinburgh, Scotland)*. 2015;24(2):93-9.
16. Skaane P et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology*. 2014;271(3):655-63.
17. McDonald ES et al. Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. *JAMA oncology*. 2016;1:737-43.
18. Miller D, Martin I, Herbison P. Interventions for relieving the pain and discomfort of screening mammography. *Cochrane Database Syst Rev*. 2002 (4):Cd002942.
19. Poulos A and Rickard M. Compression in mammography and the perception of discomfort. *Australasian radiology*. 1997;41(3):247-52.
20. Whelehan P et al. The effect of mammography pain on repeat participation in breast cancer screening: a systematic review. *Breast (Edinburgh, Scotland)*. 2013;22(4):389-94.
21. Holland K et al. Performance of Breast Cancer Screening Depends on Mammographic Compression. In: Tingberg A et al, editors. *Breast Imaging: 13th International Workshop, IWDM 2016, Malmö, Sweden, June 19-22, 2016, Proceedings*. Cham: Springer International Publishing; 2016. p. 183-9. <http://library.wur.nl/WebQuery/clc/2174165>. Cited 20/05/2017
22. Moshina N et al. Is breast compression associated with breast cancer detection and other early performance measures in a population-based breast cancer screening program? *Breast cancer research and treatment*. 2017 Mar 29.
23. Houssami N and Turner RM. Rapid review: Estimates of incremental breast cancer detection from tomosynthesis (3D-mammography) screening in women with dense breasts. *Breast (Edinburgh, Scotland)*. 2016;30:141-5.
24. Gilbert FJ et al. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. *Clinical radiology*. 2016;71(2):141-50.
25. IARC. *Breast Cancer Screening IARC Handbook of Cancer Prevention Volume 15*. ISBN: 978-92-832-3015-1. 2016.
26. European Commission Initiative on Breast Cancer: <http://ecibc.jrc.ec.europa.eu/> Cited 20/05/2017
27. Vitak B et al. Tumour characteristics and survival in patients with invasive interval breast cancer classified according to mammographic findings at the latest screening: a comparison of true interval and missed interval cancers. *European radiology*. 1999;9(3):460-9.
28. Bordas P et al. Survival from invasive breast cancer among interval cases in the mammography screening programmes of northern Sweden. *Breast (Edinburgh, Scotland)*. 2007;16(1):47-54.
29. Hofvind S et al. Mode of detection: an independent prognostic factor for women with breast cancer. *Journal of medical screening*. 2016;23(2):89-97.
30. O'Brien KM et al. Interval cancer rates in the Irish national breast screening programme. *Journal of medical screening*. 2015;22(3):136-43.
31. Fracheboud J et al. Interval cancers in the Dutch breast cancer screening programme. *British journal of cancer*. 1999;81(5):912-7.
32. Zackrisson S et al. Improved survival rate for women with interval breast cancer - results from the breast cancer screening programme in Malmö, Sweden 1976-1999. *Journal of medical screening*. 2007;14(3):138-43.
33. Hofvind S et al. 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):437-43.
34. Hoff SR et al. Breast cancer: missed interval and screening-detected cancer at full-field digital mammography and screen-film mammography-results from a retrospective review. *Radiology*. 2012;264:378-86.
35. Carbonaro LA et al. Interval breast cancers: absolute and proportional incidence and blinded review in a community mammographic screening program. *European journal of radiology*. 2014;83(2):e84-91.
36. Larsen IK et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *European journal of cancer (Oxford, England : 1990)*. 2009;45(7):1218-31.
37. Houssami N HK. The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. *Nature-npj Breast Cancer* 2017; 3: Article 12.
38. Goldhirsch A et al. 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. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2013;24(9):2206-23.
39. Williamson A, Hoggart B. Pain: a review of three commonly used pain rating scales. *Journal of clinical nursing*. 2005;14(7):798-804.
40. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*. 1990;46(4):1171-8.
41. Whitehead J. Sample size calculations for ordered categorical data. *Statistics in medicine*. 1993;12(24):2257-71.
42. Tingberg A and Zackrisson S. Digital mammography and tomosynthesis for breast cancer diagnosis. *Expert opinion on medical diagnostics*. 2011;5(6):517-26.