

Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening program (To-Be): a randomized, controlled trial

Tomosynthesis breast in screening – a randomized controlled trial

(Short title)

Hofvind, Solveig & Holen, Åsne & Aase, Hildegunn & Houssami, Nehmat & Sebuødegård, Sofie & Moger, Tron & Haldorsen, Ingrid & Akslen, Lars. (2019). Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. *The Lancet Oncology*. 20. 10.1016/S1470-2045(19)30161-5.

Summary

Background

Digital breast tomosynthesis is an advancement of mammography, and has the potential to overcome limitations of standard digital mammography (DM). This randomized, controlled clinical trial (RCT) with a parallel group design aimed to investigate the potential superiority of first generation digital breast tomosynthesis including 2D synthetic mammograms (DBT) versus DM in an organized population-based screening program.

Methods

BreastScreen Norway offers all women aged 50-69, two view (craniocaudal and medio-lateral-oblique) mammographic screening every two years and performs independent double reading with consensus. We asked all 32,976 women who attended the program in Bergen in 2016-2017 to participate in this trial. A study-specific software was developed to allocate women to either DBT or DM using a 1:1 simple randomization method based on participants' unique national identity numbers. The interviewing radiographer performed the randomization by entering the number into the software. Randomization was performed after consent and was therefore concealed from both the women and the radiographer at the time of consent; the algorithm was not disclosed to radiographers during the study period. All data needed for analyses were complete 12 months after the recruitment period ended. The primary outcome measure was screen-detected breast cancer, stratified by screening technique (DBT and DM). A log-binomial regression model was used to estimate the efficacy of DBT versus DM, defined as the crude risk ratios (RR) with 95% confidence intervals (CI) for screen-detected breast cancer for women screened during the recruitment period. A per protocol approach was used in the analyses. This RCT is registered at ClinicalTrials.gov (NCT02835625).

Findings

During the recruitment period, Jan 14, 2016 and Dec 31, 2017, 44,266 women were invited to the screening program in Bergen, and 32,976 (74.5%) attended. After excluding women with breast implants and those who did not consent to participate, 29,453 were eligible for electronic randomization. The randomization allocated 14,734 women to DBT and 14,719 to DM. Post-randomization, we excluded women with a prior breast cancer (DBT: n=314, DM: n=316), women with metastases from melanoma (DBT: n=1, DM: n=0), and those who informed the radiographer about breast symptoms after providing consent (DBT: n=39, DM: n=34). After exclusions, information from 28,749 consenting women were included in the analyses; 14,380 in the DBT arm and 14,369 in the DM arm.

The rate of screen-detected breast cancer among the screened women did not differ statistically between DBT and DM (0.66% (95/14,380) versus 0.61% (87/14,369), $p=0.56$). The RR of screen-detected breast cancer (RR=1.09, 95% CI: 0.82-1.46, $p=0.56$) did not differ statistically between the two arms.

Interpretation

This RCT indicated that DBT including synthetic 2D mammograms was as good of a screening tool as standard DM for detection of breast cancer in a population based screening program. Economic analyses and follow-up studies on interval and consecutive round screen-detected breast cancers are needed to better understand the effect of DBT in population-based breast cancer screening.

Funding

The Cancer Registry of Norway, Department of Radiology at Haukeland University Hospital, the University of Oslo, and the Research Council of Norway (the To-Be trial; project number 247941/H10 and the Centres of Excellence funding scheme; project number 223250).

Introduction

Standard digital mammography (DM) in combination with digital breast tomosynthesis (DBT) has been shown to increase the rate of screen-detected breast cancer, but its effect on recall varied in a meta-analysis of non-randomized studies (1). These studies included women of different ages, with varying screening intervals, and with a variety of screening and reading procedures (2-5). Two study designs were generally used to compare the outcome of screening with DBT+DM versus DM alone: Paired studies using women as their own controls (2-5), and unpaired studies using other geographical areas or historical data as a control group (6-9). The paired studies demonstrated a consistent increase in the detection of breast cancer, but the design hampers future analyses of interval and consecutive breast cancer. The unpaired studies demonstrated a less convincing increase in cancer detection, and the sample sizes varied considerably (1). To our knowledge, results from only one randomized controlled clinical trial (RCT) comparing DBT+DM versus DM have been published in an interim report (10). This study showed a 90% higher rate of screen-detected breast cancer for women screened with DBT+DM versus DM alone.

Lower recall rates for DBT+DM have been demonstrated, but mostly in studies from the U.S., where recall rates are known to be substantially higher than in Europe (1). A lower recall rate is beneficial for women if the rate of screen-detected breast cancer remains stable or increases. However, a higher rate of screen-detected breast cancer is only beneficial for screened women if the tumors are progressive, as opposed to small, low histologically graded tumors, which might represent overdiagnosis. If screen-detected breast cancers are detected at an earlier stage, we expect the rate of interval and/or consecutive screen-detected breast cancer to decrease as a result.

The use of DBT+DM roughly doubles the radiation dose compared to DM alone (11). As a result, synthetic 2D mammograms (SM) has been developed using raw-data from the DBT acquisition to minimize the radiation burden on women. As far as we are aware, there is insufficient evidence to draw any conclusions about the overall balance of benefits and harms of using DBT+SM in a population-based screening program, and no results from RCTs are available (11-13).

To fill the evidence gaps in the use of DBT in combination with SM in breast cancer screening, we performed a RCT in BreastScreen Norway – the Tomosynthesis trial in Bergen (the To-Be trial). This trial aimed to investigate performance measures and economic aspects of using DBT+SM (hereafter referred to as DBT) versus DM alone in an organized, population-based breast cancer screening program with a high attendance rate and complete follow-up data (14). This article presents results of outcome measures for women screened with DBT versus DM regarding the screen-detected breast cancer, consensus, recall, and distribution of histopathologic tumor characteristics, in addition to the positive predictive value of recall (PPV-1) and biopsy (PPV-2), as well as time used for screen reading.

Methods

Study design and participants

The To-Be trial was designed as a large-scale parallel group RCT where the outcome of the novel screening technique DBT was tested in an everyday screening setting. Participants were screened in Bergen through the national screening program, BreastScreen Norway, between Jan 15, 2016 and Dec 31, 2017 (recruitment period). BreastScreen Norway offers women aged 50-69 years biennial two-view mammographic screening (14). The Cancer Registry of Norway administers the program. Cancer registration is mandated by law in Norway, and the Registry's databases are >99% complete for breast cancer. This provided a unique opportunity to link the trial data with registry data on individual screening episodes, including any follow-up and diagnosis performed as a part of, or outside, BreastScreen Norway.

A total of 44,266 women born between 1947 and 1966 were invited to the screening unit in Bergen during the recruitment period (Figure 1). All attending women (n=32,976, 74.5%) were informed about the trial and received written information from an administrative assistant when they entered the screening unit. In the pre-screening room, radiographers prepared women for the screening exam and asked about participation in the trial before they asked general questions, following standard procedures. Those who agreed to participate signed an electronic consent.

The To-Be trial was approved by the Regional Committees for Medical and Health Research Ethics in South Eastern Norway (official record number 2015/424), and registered at ClinicalTrials.gov (NCT02835625).

Randomization and masking

We developed a study-specific software to randomly allocate women to receive either DBT or DM using a 1:1

simple randomization method based on participants' unique 11-digit national identity numbers. This number sequence includes two randomly generated control digits, one of which was used by the randomization software to assign the screening technique. The interviewing radiographer in the pre-screening room performed the randomization by entering the unique identification number of the consenting women into the software. Randomization was done after consent was obtained and was therefore concealed from both the women and the radiographer at the time of consent. Only the PI and the software developers knew the algorithm for randomization; it was concealed to the radiographers during the study period. We evaluated the randomization process continuously, but did not evaluate the masking of the trial arm allocation. Due to the characteristics of the intervention, it was not possible to blind the intervention neither in the screening nor in the reading process.

Attendance rates in the program and the trial were accessible for all professionals involved throughout the duration of the study. No other results were made available during the recruitment period. Quality control and assurance was continuously performed according to standard procedures, as interim analyses, to ensure the women's safety according to radiation dose, recall and breast cancer detection, and for continuation of the trial. In the case of unacceptable values, the trial would have been discontinued. Results were regularly presented for the Steering Committee of the trial. Results were only discussed in the Committee. No independent review was performed.

Procedures

Consenting women underwent one screening examination during the recruitment period. A standard examination consisted of two-view (craniocaudal and mediolateral oblique views) mammography of each breast, either with DBT including SM or DM, using imaging equipment from GE Health Care (SenoClaire 3D Breast Tomosynthesis™). Two radiographers performed the screening examination as a team and all examinations took place at one of the two equally equipped examination rooms. The DBT acquisition consisted of nine exposures over an angle of 25°, reconstructed into SM, 10mm slabs, and 1mm planes. Mean glandular dose per exam was 2.96 mGy for DBT and 2.95 mGy for DM during the first year of the trial (15). Independent double reading was performed on IDI workstations, each with two 5-megapixel monitors (GE Healthcare MammoWorkstation Version 4.7.0 Image Diagnost).

A standardized reading protocol including two-view SM, slabs and planes of each breast for DBT or two-view DM of each breast, was used for screen reading (Appendix A) and at the consensus meetings (hereafter referred as consensus). Up to four prior screening examinations, and/or diagnostic images from the previous ten years were available on the workstation for both trial arms, both for screen reading and at consensus.

We performed independent double-reading with consensus, following the program's standard procedures (14). Each breast was assigned a score of 1-5 from each radiologist. A score of 1 indicated the screening examination was negative for abnormality; 2, probably benign; 3, intermediate suspicion; 4, probably malignant; and 5, high suspicion of malignancy. If either radiologist assigned a score of 2 or higher, consensus was used to determine whether to recall the woman for further assessment, hereafter referred to as recall. Consensus was performed by pairs of radiologists, and a third radiologist was consulted if the pair could not agree.

A pool of eight breast radiologists performed the initial screen readings and consensuses. Their experience in screen reading (screen-film and DM) prior to start-up of the RCT varied from zero to roughly 110,000 examinations (15). The number of DBT and DM screen readings, and interpretation time in the trial were automatically recorded for each radiologist, while the consensus time was recorded for each meeting.

Further assessment took place 2-8 weeks after screening, at the breast center at Haukeland University Hospital by the same pool of radiologists who did the screen reading. If indicated, additional imaging, including DBT, ultrasound, and less frequently contrast enhanced spectral mammography, and/or MRI, was performed. Cancer diagnoses were verified by pathologists examining histological specimens. Information about histopathologic features were considered complete 12 months after further assessment took place.

Women were the unit of analysis and since all women were screened only once in the trial, the number of women and screening examinations were the same. Women with bilateral breast cancer were included once based on the topography of the recalled lesion or, if recalled for lesions in both breasts, based on malignancy according to histologic type (invasive before ductal carcinoma in situ, DCIS), tumor diameter and histologic grade. Histopathologic tumor characteristics for DCIS and invasive breast cancer were determined from routine histopathology reports performed by pathologists affiliated with the breast center at Haukeland University Hospital, with no reclassification. Lobular carcinoma in situ (LCIS) was included in the group of benign lesions (n=3 for DBT and n=2 for DM) (16). We diagnosed one case of pleomorphic lobular carcinoma in situ in the DBT arm, which we included in the LCIS group.

We collected information about the screening examinations and screen reading electronically, in compliance with standard procedures at BreastScreen Norway. Additional information related to consensus and recall was manually recorded on a paperbased form, which was designed specifically for this trial by the PI and the radiologists at the breast center. These forms were transferred electronically or by letter mail to the Cancer Registry of Norway, where the data were registered and quality assured by a dedicated research assistant before interim and final analyses were performed. All study data were stored in databases at the Cancer Registry.

Outcomes

The main objective of this RCT was to determine whether the rate of screen-detected breast cancer was favorable for DBT versus DM, as pre-specified in the protocol. Breast cancer was defined as histologically verified DCIS and/or invasive breast cancer.

Secondary outcome measures, also specified in the protocol, were recalls, positive predictive value of recall (PPV-1) and biopsy (PPV-2), histopathologic tumor characteristics, and economic aspects. PPV-1 was the percentage of breast cancer cases detected among those recalled and PPV-2 the percentage of breast cancer detected among recalled women who underwent a needle biopsy.

Prognostic characteristics for invasive tumors included mean and median tumor diameter, and distribution of tumor diameter groups (<10 mm, 10-<20 mm and \geq 20 mm), histologic grade (1, 2 and 3), and lymph node involvement (negative/positive) were presented as percentages of all values. For tumor diameter, lesions treated with neoadjuvant therapy were included in the category “information not available” in the respective tables (n=9 for DBT and n=11 for DM). Predictive biomarkers included estrogen and progesterone receptor status (ER+/- and PR +/-), human epidermal growth factor 2 receptor status (Her2 +/-), and Ki67 proliferation (<30% and \geq 30%). This information was used to classify the invasive tumors into subtypes (Luminal A, Luminal B Her2-, Luminal B Her2+, Her2+, and Triple negative)(17).

Other outcomes, as consensus, time spent on screen reading and consensus, mammographic features and radiation doses for the two techniques were also included in the protocol, as well as interval and breast cancers in consecutive screening round. At least two years of follow-up of each individual woman is needed to obtain complete data for latter two outcomes.

We defined women’s screening history as prevalent or subsequent, the former indicating the first or incident screen in BreastScreen Norway, and the latter indicating prior attendance in the program and the availability of prior screening mammograms (DM) for comparison during screen reading.

Statistical analysis

The hypothesis of our RCT was that the rate of screen-detected breast cancer would be superior for DBT versus DM. In a population with an estimated prior screen-detected breast cancer rate of 0.60%, we calculated that with 15,000 women in each arm, we could observe an detection from 0.60% with DM to 0.88% with DBT, given 80% power and a two-sided significance threshold of 5%.

Our analyses included information from all women with a complete screening examination, who had no prior history of breast cancer or metastatic melanoma, and who did not report breast symptoms when attending for screening examination. These women represent the per-protocol population. No women revoked their consent in the study. Results from interim analyses on consensus and recall rate, in addition to radiation dose for DBT versus DM from the first year of the recruitment period were performed as planned (15).

Our primary goal was to estimate crude risk ratios (RR) with 95% confidence intervals (CI) for screen-detected breast cancer for DBT with DM as the reference, using a log-binomial regression model. Log-binomial regression models were also fitted to estimate the RR of recall for women screened with DBT versus DM.

All analyses were stratified by the randomly assigned screening technique (DBT or DM). Age was categorized into four groups; <55, 55-59, 60-64 and \geq 65 years. Consensus, recall, biopsy, and screen-detected breast cancer were presented as rates per 100 screened women within the recruitment period, while PPV-1 and 2 were presented as percentages. The distributions of histopathologic tumor characteristics were reported as percentages for cases with non-missing values. Tumor diameter (mm) and time spent on initial and consensus reading (min:sec) did not follow a perfect normal distribution, thus we estimated both mean with standard deviation (SD), and median values with interquartile range (IQR).

For sensitivity analyses, we performed log-binomial regression models to estimate crude RR for screen-detected breast cancer for DBT versus DM, stratified by screening history.

We used STATA version 15.0 (Stata Corp, TX) for all statistical analyses and tested differences across categories using two-sample t-tests, chi-square tests, one-way ANOVA or tests of proportions (Z-test). A p-value of <0.05 was considered statistically significant.

Role of funding source

The funder of the study had no role in the study design, data collection, analyses, interpretation, or writing of the report. SH, ÅH, and SS had full access to all the data in the study, and the corresponding author (SH) had final responsibility for the decision to submit the manuscript.

Results

Among the 44,266 women invited to the screening unit in Bergen during the recruitment period, Jan 15, 2016 – Dec 31, 2017, 32,976 (74.5 %) attended (Figure 1). Non-consenting women (9.1 %, 2999/32,976) and those with breast implants (1.6%, 524/32,976) were not included in the trial, leaving 29,453 (89.3 %) eligible for randomization. Of those, 14,734 were allocated to screening with DBT and 14,719 to DM. Additional exclusions post-randomization (breast cancer diagnosed before date of screening, metastatic melanoma, and reporting symptoms of breast cancer at the screening examination) resulted in a final study population of 28,749 women; 14,380 randomly assigned to DBT and 14,369 to DM during the recruitment period (Figure 1). Baseline characteristics for the study group are presented in Table 1 while characteristics prior to post-randomization exclusions are available in Appendix B.

The number of women diagnosed with screen-detected breast cancer due to mammographic findings was 182; 31 DCIS and 151 invasive cancers (Table 2). Four women were diagnosed with bilateral cancer (n=2 for DBT and n=2 for DM). The rate of screen-detected breast cancer did not differ between DBT and DM 0.66 % (95/14,380) versus 0.61 % (87/14,369, p=0.56), respectively. The risk ratio of screen-detected breast cancer did not differ between the two techniques (RR=1.09, 95 % CI: 0.82-1.46, p=0.56), while the risk of recall was lower for DBT versus DM (RR=0.78, 95 % CI: 0.69-0.88, p<0.0001). Summary of outcomes, without any post-randomization exclusions, are available in Appendix C.

Consensus was 6.3 % (908/14,380) for DBT and 7.4 % (1060/14,369) for DM (p=0.0004) (Table 2). Recall was 3.1 % (444/14,380) for DBT and 4.0 % (571/14,369) for DM (p<0.0001). The biopsy rate did not differ between DBT and DM (1.8% (252/14,380) versus 1.9 % (271/14,369), p=0.40, respectively). PPV-1 was statistically higher for DBT (21.4 %; 95/444) compared to DM (15.2 %; 87/571, p=0.01), while PPV-2 was not (p=0.18). RR for recall and screen-detected breast cancer without any post-randomization exclusions and results of sensitivity analyses, stratified by screening history are available in Appendix D and E, respectively.

Consensus, recall and detection increased during the study period for DBT (p=0.02, 0.01, and 0.04, respectively) (Figure 2a, b and c). For DM, an increase was observed for recall (p=0.01).

The distribution of histopathological tumor characteristics did not differ statistically between the two arms of the RCT for invasive breast cancer (Table 3), or for DCIS (Table 4).

All eight radiologists read both DBT and DM during the trial, ranging from 1079 to 5663 examinations for DBT and from 1067 to 7538 for DM, per radiologist (Appendix F). Overall, mean time spent on initial screen reading was 01:06 (median 00:48, IQR: 00:45) for DBT, including interpretation of SM and DBT planes as well as priors, and 00:39 (median 00:23, IQR: 00:31) for DM, including interpretation of priors (p for difference in reader time between DBT and DM: <0.0001). Mean time for each radiologist ranged from 00:39 to 02:42 for DBT and from 00:13 to 03:02 for DM. Mean time spent on consensus was 02:51 with SD 01:48 (median: 02:21, IQR:01:50) for DBT and 02:04 with SD 02:05 (median: 01:42, IQR: 1:11) for DM (p<0.0001), respectively (Appendix G).

Discussion

This large-scale RCT compared results from screening with DBT including SM, with standard DM, in an organized population-based breast cancer screening program. The breast cancer detection did not differ statistically significantly for women screened with DBT versus DM. We observed lower consensus and recall for women screened with DBT versus DM, while the PPV for recalls was higher for DBT compared to DM. The distribution of histopathologic tumor characteristics did not differ statistically significantly between the two screening techniques.

Our finding of no statistical difference in breast cancer detection for DBT versus DM was inconsistent with results from the majority of both paired and unpaired studies (1). Our intervention arm used DBT+SM, whereas most of the others studies have used DBT+DM. A similar effect of SM and DM in combination with DBT has

been consistently reported (3, 18-20), however, the quality of SM may present differently for different DBT machines and software versions.

It is commonly assumed that DM is superior to DBT in the characterization of microcalcifications. However, recent studies have shown that the perceptibility of microcalcifications is also adequate for DBT in combination with SM (21, 22). Several factors influence image quality of both DBT, including SM, and DM, such as filter/anode combinations, spatial resolution, the angular range of the x-ray tube and radiation dose (21, 23). It could be argued that the radiation dose measured for DBT in our study, which was lower than that reported in other studies of DBT (15), may have negatively affected the image quality. However, differences in vendor-specific technical implementations and the optimization of mammography workstations can also affect image quality. Moreover, different requirements for training to start screen reading, reading conditions and protocols, including the availability and use of prior mammograms, could have influenced the radiologists' perception and interpretation of mammographic features, and thus contributed to the heterogeneity of results in the published literature to date.

We used two-view mammography with independent double-reading and separate consensus in both trial arms and a hanging protocol with availability of screening and any diagnostic mammograms taken during the previous ten years, in line with standard procedures. However, the radiologists' experience in DM screen reading prior the start of the trial varied, and the program's recommendations of 5000 annual screen readings was met by half of the radiologists (15). Further, reading volume varied between radiologists for DBT and DM during the trial. It has been claimed that experience in screen reading and preferences of DM or DBT might influence reader sensitivity (24, 25). Radiologists in this study were not exposed to any preliminary results during any part of the trial, which we consider a strength; this made it impossible to provide individual feedback about their performance. Nonetheless, the radiologists were involved in both screen reading and recall assessments and thus not blinded to the final outcome of the women's screening examination.

All participating radiologists were somewhat trained in DBT screen reading and diagnostics prior to the start of the trial, but they did not practice it in an everyday screening setting until the trial started (15). A lack of experience in screen reading of DBT is therefore considered a limitation of our study (26) and it is possible that the radiologists had not yet achieved optimal screen reading capabilities with DBT at the start-up of the trial.

Despite this potential limitation, time used for screen reading was lower in the To-Be trial than reported from the Oslo Tomosynthesis Screening Trial, 66 seconds for DBT and 39 seconds for DM in our study, compared with 91 seconds for DBT+DM and 45 seconds for DM alone in the Oslo trial (4). This could represent cultural differences between the two breast centers, although our results showed substantial variation in time used for screen reading between the radiologists.

Analysis of the radiologist's sensitivity, interval and consecutive round screen-detected breast cancer, and reviews of cases dismissed at consensus, interval and next round screen-detected breast cancers are needed to conclude whether the cancer cases were missed due to interpretation errors. Admittedly, consensus, recall and detection increased over the two years recruitment period for DBT, while an increase was observed only for recalls for DM. These findings might represent a learning effect for DBT. The increase in recall for DM might represent natural variation, or study effect.

The availability of prior screening and diagnostic images at screen reading might have influenced the number of cases discussed at consensus, and cases dismissed at consensus. It is possible that findings on DBT images were present and perceptible by radiologists on the prior DM, but considered non-suspicious and interpreted as negative, or positive but dismissed on the consensus. Lack of obvious findings on the current SM and/or the slabs and planes might have downgraded their interpretation or dismissed cases at consensus (27). Furthermore, the availability of several sets of prior mammograms could have been distracting instead of elucidative to the radiologists. These factors might have influenced the rates of recall and breast cancer detection, as well as the histopathological tumor characteristics in the DBT arm. We presume that calling women with these findings back for further assessment might have led to a higher detection of breast cancer in our study, however, we do not know whether these lesions represent small tumors of low histologic grade or "killing cancers".

Our results on histopathologic tumor characteristics are in line with those from the Malmö tomosynthesis trial (5). The distribution of tumor diameter, histologic grade and tumor cell proliferation index by Ki67 staining indicated an increase in the detection of node-negative progressive tumors, which could be considered favorable for DBT. The relatively high number of cases with no available tumor diameter may have affected the mean tumor diameter in our study and is likely related to new guidelines that recommend neoadjuvant treatment for some tumors ≥ 20 mm (28). Our histopathologic findings diverge somewhat from some of the other studies, which have shown an increase in small, low histologic grade breast cancers in groups screened with DBT (18, 29).

To investigate a possible study effect, we compared recall in Bergen prior to the trial, 2008-2015, with results from the trial. The recall prior to the trial was 3.3%, which did not differ from the recall for DBT (3.1%, $p=0.18$), in contrast to the recall for DM which was higher (4.0%, $p<0.0001$). The change in recall in Bergen is likely a study effect and is expected to affect both arms equally. However, our study was designed to compare results of DBT versus DM regardless of this underlying effect. Recall for Norway, Bergen excluded, was 2.9% in the pre-trial period, while it increased to 3.3% during the trial period, which indicate a higher recall in Bergen than the rest of Norway, irrespective of the trial. The breast cancer detection in our trial did not differ from the pre-trial period (0.62%), either for DBT (0.66%, $p=0.60$) or for DM (0.61%, $p=0.79$). Similar comparison for BreastScreen Norway, Bergen excluded, showed a screen-detected breast cancer of 0.55% in the pre-trial period and 0.59% in the trial period. The pre-trial rates did not differ statistically from the rates shown in the trial in either Bergen or in Norway. These comparisons are not consistent with a possible study effect on the primary outcome of the trial.

Despite several studies demonstrating a higher rate of screen-detected breast cancer with DBT than DM, the implementation of DBT in population-based screening programs has been anything but rapid. This could be due to lack of evidence from studies with sufficiently large study samples and robust study designs, as well as the problematic finding of an increased incidence of small, low histologic grade tumors without a corresponding decrease in the rate of interval breast cancers. Given the results described above regarding rates of screen-detected breast cancer for DBT, and since our trial is one of the earliest program-embedded RCTs of DBT screening, we present a reflection on factors that might have shaped our study findings in order to inform future breast cancer screening research.

Running a RCT of new technology in an everyday screening setting is a challenging task, which the To-Be trial managed proficiently by masking interim results, using a closed pool of radiographers and radiologists, and by using standardized imaging, screen reading and consensus procedures in both arms; our findings therefore reflect real-world screening outcomes for a RCT.

The population-based RCT design, a high participation rate, and a high level of data completeness due to linkage with the Cancer Registry of Norway are all strengths of the To-Be trial. However, there are some limitations to our trial. First, our assumed rate of screen-detected breast cancer in the DBT arm (based on current knowledge at study inception) was somewhat exaggerated, leading to diminished statistical power. In retrospect, a combination of a superiority and non-inferiority RCT might have been a better design (30), given the observed detection rate and the final study sample size. For the observed breast cancer rates, we would have needed about 400,000 women in each arm to detect statistically significant difference.

Increasing the sample size in our RCT would have been difficult, as an extension of study period would have resulted in the inclusion of women with DBT as a prior screening exam, instead of increasing the sample size of women first time screened with DBT. Since we included all women who attended the screening unit in Bergen during the recruitment period, a multicenter study would be the only way to increase the study population.

Another limitation is that the To-Be trial is a single-center trial using equipment from one vendor. Our results may therefore have limited generalizability to other settings. Additionally, we do not currently have sufficiently long follow-up period to report on interval breast cancers or breast cancer mortality. A longer follow-up time is also needed for analyses on patient reported outcome measures, which is on our long-term plan. Further, we expect to stratify our findings by mammographic density in later analyses.

Lastly, although we did several analyses, we tested and found that employing a statistical correction for multiple testing, such as the Bonferroni correction would not alter our conclusions.

In summary, our population-based breast cancer screening RCT using first generation DBT with SM versus DM alone did not identify a statistical difference in the rate of screen-detected breast cancer. The distribution of histopathologic tumor characteristics did not differ between the two arms either. A lower recall rate and a higher positive predictive value for recall was found for women screened with DBT compared to those screened with DM. Our results indicate that use of DBT in a screening setting is safe for women at average risk of breast cancer. Further studies with longer follow-up that analyze interval and screen-detected breast cancer in the consecutive screening round are needed to better understand the effect of DBT in a population-based screening setting.

References

1. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. *Journal of the National Cancer Institute* 2018; 110: 942-9.

2. Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *The Lancet Oncology* 2013; 14: 583-9.
3. Bernardi D, Macaskill P, Pellegrini M, Valentini M, Fanto C, Ostilio L, 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: 1105-13.
4. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology* 2013; 267: 47-56.
5. Zackrisson S, Lang K, Rosso A, Johnson K, Dustler M, Fornvik D, et al. One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. *The Lancet Oncology* 2018; 19: 1493-503.
6. Conant EF, Beaber EF, Sprague BL, Herschorn SD, Weaver DL, Onega T, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. *Breast cancer research and treatment* 2016; 156: 109-16.
7. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR American journal of roentgenology* 2014; 203: 687-93.
8. Friedewald SM, Rafferty EA, Conant EF. Breast cancer screening with tomosynthesis and digital mammography-reply. *Jama* 2014; 312: 1695-6.
9. McCarthy AM, Kontos D, Synnestvedt M, Tan KS, Heitjan DF, Schnall M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *Journal of the National Cancer Institute* 2014; 106(11).
10. Pattacini P, Nitrosi A, Giorgi Rossi P, Iotti V, Ginocchi V, Ravaioli S, et al. Digital Mammography versus Digital Mammography Plus Tomosynthesis for Breast Cancer Screening: The Reggio Emilia Tomosynthesis Randomized Trial. *Radiology* 2018; 288: 375-85.
11. Houssami N. Evidence on Synthesized Two-dimensional Mammography Versus Digital Mammography When Using Tomosynthesis (Three-dimensional Mammography) for Population Breast Cancer Screening. *Clinical breast cancer* 2017; 18: 255-60.
12. Wahab RA, Lee SJ, Zhang B, Sobel L, Mahoney MC. A comparison of full-field digital mammograms versus 2D synthesized mammograms for detection of microcalcifications on screening. *European journal of radiology* 2018; 107: 14-9.
13. Bernardi D, Li T, Pellegrini M, Macaskill P, Valentini M, Fanto C, et al. Effect of integrating digital breast tomosynthesis (3D-mammography) with acquired or synthetic 2D-mammography on radiologists' true-positive and false-positive detection in a population screening trial: A descriptive study. *European journal of radiology* 2018;106:26-31.
14. Hofvind S, Tsuruda K, Mangerud G, Ertzaas AKO, Holen AS, Pedersen K et al. The Norwegian Breast Cancer Screening Program, 1996-2016: Celebrating 20 years of organised mammographic screening. In: *Cancer in Norway 2016 - Cancer incidence, mortality, survival and prevalence in Norway*. Oslo: Cancer Registry of Norway, 2017. ISBN 978-82-473-0055-8.
15. Aase HS, Holen AS, Pedersen K, Houssami N, Haldorsen IS, Sebuodegard S, et al. A randomized controlled trial of digital breast tomosynthesis versus digital mammography in population-based screening in Bergen: interim analysis of performance indicators from the To-Be trial. *European radiology* 2018. Epub ahead of print
16. *AJCC Cancer Staging Manual*. Springer International Publishing, Eith Edition, 2017, ISBN: 978-3-319-40617-6.
17. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thurlimann B, 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: 2206-23.

18. Skaane P, Bandos AI, Eben EB, Jepsen IN, Krager M, Haakenaasen U, 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: 655-63.
19. Hofvind S, Hovda T, Holen AS, Lee CI, Albertsen J, Bjørndal H, et al. Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. *Radiology* 2018; 287: 787-94.
20. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical Performance of Synthesized Two-dimensional Mammography Combined with Tomosynthesis in a Large Screening Population. *Radiology* 2017; 283: 70-6.
21. Mackenzie A, Kaur S, Elangovan P, Dance DR, Young KC, editors. Comparison of synthetic 2D images with planar and tomosynthesis imaging of the breast using a virtual clinical trial. *SPIE Medical Imaging*; 2018: SPIE; 2018. p. 9.
22. Rodriguez-Ruiz A, Lardenoije S, Wanders AJT, Sechopoulos I, Mann RM. Comparison of breast cancer detection and depiction between planar and rotating synthetic mammography generated from breast tomosynthesis. *European journal of radiology* 2018; 108: 78-83.
23. Rodriguez-Ruiz A, Teuwen J, Vreemann S, et al. New reconstruction algorithm for digital breast tomosynthesis: better image quality for humans and computers. *Acta radiologica (Stockholm, Sweden:1987)* 2018; 59: 1051-9.
24. Tucker L, Gilbert FJ, Astley SM, Dibden A, Seth A, Morel J, et al. Does Reader Performance with Digital Breast Tomosynthesis Vary according to Experience with Two-dimensional Mammography? *Radiology* 2017; 283: 371-80.
25. Alakhras MM, Brennan PC, Rickard M, Bourne R, Mello-Thoms C. Effect of radiologists' experience on breast cancer detection and localization using digital breast tomosynthesis. *European radiology* 2015; 25: 402-9.
26. Skaane P, Bandos A, Niklason L, Sebuøegård S, Østerås B, Gullien R, et al. Digital Mammography versus Digital Mammography plus Tomosynthesis in Breast Cancer Screening: The Oslo Tomosynthesis Screening Trial. *Radiology*, In press. 2019.
27. Iotti V, Rossi PG, Vacondi R, Nitrosi A, Bertolini M, Bacchini E et al. Comparing Two Visualization Protocols for Tomosynthesis in Screening: Specificity and Sensitivity of Slabs vs Slabs plus Planes. Abstract RSNA, 2017. [cited 2019 Jan 30]. Available from <https://ecronline.mysr.org/ecr2018/index.php?p=recording&t=recorded&lecture=comparing-two-visualisation-protocols-for-tomosynthesis-in-screening-specificity-and-sensitivity-of-slabs-vs-planes-plus-slabs>
28. Curigliano G, Burstein HJ, E PW, Gnant M, Dubsky P, Loibl S, et al. De-escalating and escalating treatments for early-stage breast cancer: the St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2017; 28: 1700-12.
29. Caumo F, Romanucci G, Hunter K, Zorzi M, Brunelli S, Macaskill P, et al. Comparison of breast cancers detected in the Verona screening program following transition to digital breast tomosynthesis screening with cancers detected at digital mammography screening. *Breast cancer research and treatment* 2018; 170: 391-7.
30. Snapinn SM. Noninferiority trials. *Current controlled trials in cardiovascular medicine* 2000; 1: 19-21

Figure legends

Figure 1: Study population in the To-Be trial

Figure 2: Consensus (upper panel), recall (middle panel) and screen-detected breast cancer (SDC, lower panel) per 100 women screened, by time since start-up of the To-Be trial, and screening technique (DBT and DM). Tested for trend with one-way ANOVA.

Research in Context

Evidence before this study

Two of the investigators (SH and NH) undertook a literature search (MEDLINE: 2010 to August 2015; explored 'breast neoplasms' and searched 'tomosyn\$' in the title) to identify studies reporting screening outcomes for population-based breast screening with digital breast tomosynthesis (DBT): there were no RCTs of DBT screening. Three prospective non-randomized trials and several retrospective studies reported screening performance measures for DBT, predominantly in combination with standard digital mammography (DM). This indicated that DBT+DM generally improved screening outcome measures by increasing cancer detection and/or reducing recall compared to DM alone, although effect estimates varied. Only two of these studies investigated screen reading using DBT alone, or investigated DBT including synthetic 2D mammograms (SM) (as one of the screen reading strategies). Both studies were prospective and reported only interim analyses. The former showed that DBT alone significantly increased cancer detection and recall rate, while the latter showed equivalence in cancer detection between DBT+SM and DBT+DM.

Added value of this study

To our knowledge, this is the first RCT reporting outcomes for screening with DBT+SM in a high-throughput, population-based breast cancer screening program. In contrast to the majority of other studies, we did not find statistically significant differences in breast cancer detection rates or tumor characteristics for DBT+SM versus DM alone. We found a lower recall rate and higher positive predictive value of recalls among those screened with DBT compared to DM. Our results indicate that DBT+SM is as good a screening tool as standard DM for selected breast cancer screening outcome measures in a population based screening program for women at average risk of breast cancer. These results will be valuable in the coming policy discussions about whether to implement DBT in breast cancer screening programs.

Implication of all the available evidence

The evidence on DBT as a screening tool points towards a higher rate of screen-detected breast cancer and prognostically favorable tumor characteristics among women screened with DBT in combination with DM or SM, compared with standard DM. In our study, we compared outcome measures for DBT+SM versus standard DM in a modern screening program, BreastScreen Norway, which offers women aged 50-69 biennial mammography screening, and independent double reading with consensus. Results from our real-world, screening-based RCT using DBT+SM as the intervention, will inform upcoming policy decisions about implementing DBT in an organized, population-based screening program for breast cancer. Our results indicate that use of DBT in a screening setting is safe, and has the potential to reduce harms associated with mammographic screening. Further studies on interval and screen-detected breast cancer in consecutive screening rounds are needed to better understand the effect of DBT in a population-based screening setting. Additionally, more knowledge is needed about the recommendations and, ultimately, the requirements for radiologists to start screen reading with DBT, as well as the economic aspects of implementing DBT in a screening setting. These knowledge gaps should be filled, and our results should be replicated in other national screening programs, in order to make an evidence-based decision regarding whether DBT should replace DM as the standard tool for breast cancer screening.

Authors' contributions

Guarantors of integrity of entire study: All authors

Study concepts/study design or data acquisition or data analysis/interpretation: All authors

Manuscript drafting or manuscript revision for important intellectual content: SH, ÅSH, NH, SS, LAA

Manuscript editing: All authors.

Approval of final version of submitted manuscript: All authors

Agrees to ensure any questions related to the work are appropriately resolved: All authors

Literature research: SH, ÅSH, HSAA, NH

Statistical analysis: SH, ÅSH, NH, SS, LAA

Declaration of interest

NH was supported by a grant from the National Breast Cancer Foundation Cancer Research Leadership Fellowship. SH is the head of BreastScreen Norway. All other authors declare no competing interests.

Acknowledgement

We want to thank the radiographers and the radiologists at the breast centre at the Department of radiology and at the breast center at Haukeland University Hospital, those who screened the women and did the screen reading,

for their support and enthusiasm for the study. Thanks to Anders S Danielsen and Kaitlyn Tsuruda at the Cancer Registry for help during the revision of the manuscript.

Table 1: Baseline characteristics of the women included in the To-Be trial

	Screening technique					
	Total n = 28,749		DBT n = 14,380		DM n = 14,369	
	n	% (95% CI)	n	% (95% CI)	n	% (95% CI)
Age						
<55 years	7559	26.3 (25.8-26.8)	3746	26.1 (25.3-26.8)	3813	26.5 (25.8-27.3)
55-59 years	7296	25.4 (24.9-25.9)	3628	25.2 (24.5-25.9)	3668	25.5 (24.8-26.2)
60-64 years	7117	24.8 (24.3-25.3)	3625	25.2 (24.5-25.9)	3492	24.3 (23.6-25.0)
>64 years	6777	23.6 (23.1-24.1)	3381	23.5(22.8-24.2)	3396	23.6 (22.9-24.3)
Screening history						
Prevalent	4066	14.1 (13.7-14.6)	2013	14.0 (13.4-14.6)	2053	14.3 (13.7-14.9)
Subsequent	24,683	85.9 (85.4-86.3)	12,367	86.0 (85.4-86.6)	12,316	85.7 (85.1-86.3)

Table 2: Summary of outcomes in the To-Be trial

	Screening technique						p-value [‡]
	Total n = 28,749		DBT n = 14,380		DM n = 14,369		
	n	% (95% CI)	n	% (95% CI)	n	% (95% CI)	
SDC [#]	182	0.63 (0.54-0.73)	95	0.66 (0.53-0.79)	87	0.61 (0.48-0.73)	0.56
DCIS [‡]	31	0.11 (0.07-0.15)	15	0.10 (0.05-0.16)	16	0.11 (0.06-0.17)	0.86
Invasive	151	0.53 (0.44-0.62)	80	0.56 (0.43-0.68)	71	0.49 (0.38-0.61)	0.47
Recall	1015	3.5 (3.3-3.8)	444	3.1 (2.8-3.4)	571	4.0 (3.7-4.3)	<0.0001
PPV-1		17.9 (15.6-20.4)		21.4 (17.6-25.2)		15.2 (12.3-18.2)	0.01
PPV-2		34.8 (30.7-39.1)		37.7 (31.7-43.7)		32.1 (26.5-37.7)	0.18
Biopsy	523	1.8 (1.7-2.0)	252	1.8 (1.5-2.0)	271	1.9 (1.7-2.1)	0.40
Consensus	1968	6.8 (6.6-7.1)	908	6.3 (5.9-6.7)	1060	7.4 (6.9-7.8)	0.0004

[#]SDC: Screen-detected breast cancer

[‡]DCIS: Ductal Carcinoma In Situ

[‡] p-value for Z-test between DBT and DM

Table 3: Histopathologic tumor characteristics of invasive screen-detected breast cancer

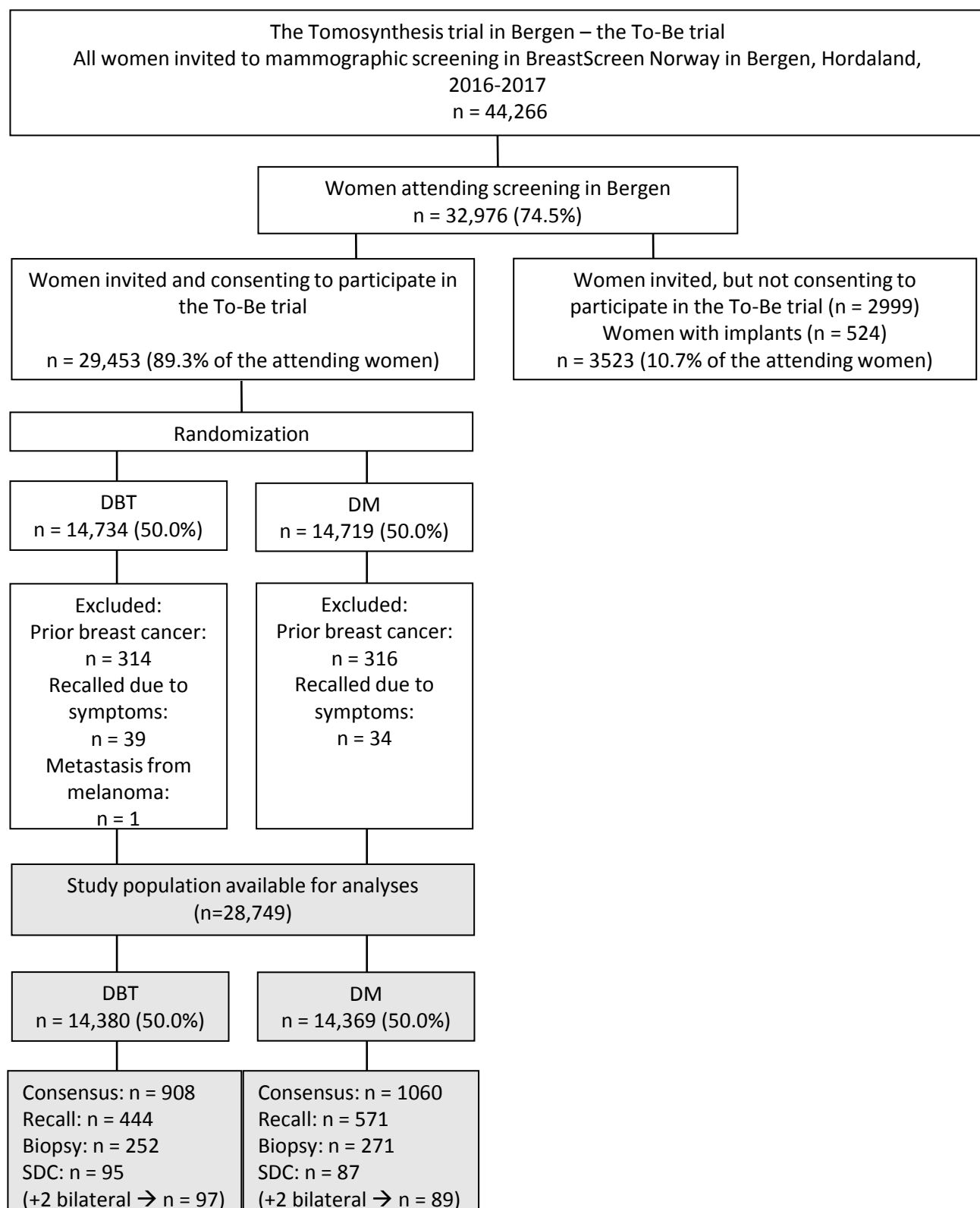
	Total n = 151		DBT n = 80		DM n = 71		p-value
	n	% (95% CI)	n	% (95% CI)	n	% (95% CI)	
Histologic type							
Carcinoma NST	113	74.8 (67.1-81.5)	62	77.5 (66.8-86.1)	51	71.8 (59.9-81.9)	0.11 [‡]
Lobular carcinoma	19	12.6 (7.7-19.0)	6	7.5 (2.8-15.6)	13	18.3 (10.1-29.3)	
Tubular carcinoma	6	4.0 (1.5-8.4)	2	2.5 (0.3-8.7)	4	5.6 (1.6-13.8)	
Other carcinoma [‡]	13	8.6 (4.7-14.3)	10	12.5 (6.2-21.8)	3	4.2 (0.9-11.9)	
Tumor diameter [§]							
Mean, SD (mm)	130	15.3, 8.5	70	16.0, 8.4	60	14.5, 8.8	0.33 [€]
Median, IQR (mm)	130	14.0, 8.6	70	14.9, 7.0	60	14.0, 10.3	
<10 mm	28	21.5 (14.8-29.6)	10	14.3 (7.1-24.7)	18	30.0 (18.8-43.2)	0.09 [‡]
≥10-<20 mm	72	55.4 (46.4-64.1)	43	61.4 (49.0-72.8)	29	48.3 (35.2-61.6)	
≥20 mm	30	23.1 (16.1-31.3)	17	24.3 (14.8-36.0)	13	21.7 (12.1-34.2)	
Information not available	21		10		11		
Lymph node positive	32	21.5 (15.2-28.9)	14	17.7 (10.0-27.9)	18	25.7 (16.0-37.6)	0.24 [‡]
Information not available	2		1		1		
Histologic grade							
1	46	31.7 (24.3-40.0)	22	28.9 (19.1-40.5)	24	34.8 (23.7-47.2)	0.53 [‡]
2	73	50.3 (41.9-58.7)	38	50.0 (38.3-61.7)	35	50.7 (38.4-63.0)	
3	26	17.9 (12.1-25.2)	16	21.1 (12.5-31.9)	10	14.5 (7.2-25.0)	
Information not available	6		4		2		
ER+	140	93.3 (88.1-96.8)	71	89.9 (81.0-95.5)	69	97.2 (90.2-99.7)	0.07 [‡]
Information not available	1		1		0		
PR+	123	82.0 (74.9-87.8)	61	77.2 (66.4-85.9)	62	87.3 (77.3-94.0)	0.11 [‡]
Information not available	1		1		0		
Her2+	16	10.7 (6.2-16.7)	8	10.1 (4.5-19.0)	8	11.3 (5.0-21.0)	0.82 [‡]
Information not available	1		1		0		
Ki67 ≥30%	34	25.0 (18.0-33.1)	21	29.6 (19.3-41.6)	13	20.0 (11.1-31.8)	0.20 [‡]
Information not available	15		9		6		
Subtype							
Luminal A	87	60.0 (51.5-68.0)	44	58.7 (46.7-69.9)	43	61.4 (49.0-72.8)	0.43 [‡]
Luminal B Her2-	36	24.8 (18.0-32.7)	18	24.0 (14.9-35.3)	18	25.7 (16.0-37.6)	
Luminal B Her2+	12	8.3 (4.3-14.0)	5	6.7 (2.2-14.9)	7	10.0 (4.1-19.5)	
Her2+	4	2.8 (0.8-6.9)	3	4.0 (0.8-11.2)	1	1.4 (0.04-7.7)	
Triple Negative	6	4.1 (1.5-8.8)	5	6.7 (2.2-14.9)	1	1.4 (0.04-7.7)	
Information not available	6		5		1		

[‡] Including mucinous and other invasive cancers[§] Neo-adjuvant treated women had no information available about tumor diameter[‡] p-value for Chi-square test between DBT and DM[€] p-value for t-test of mean tumor diameter, between DBT and DM**Table 4: Histopathologic tumor characteristics of screen-detected ductal carcinoma in situ**

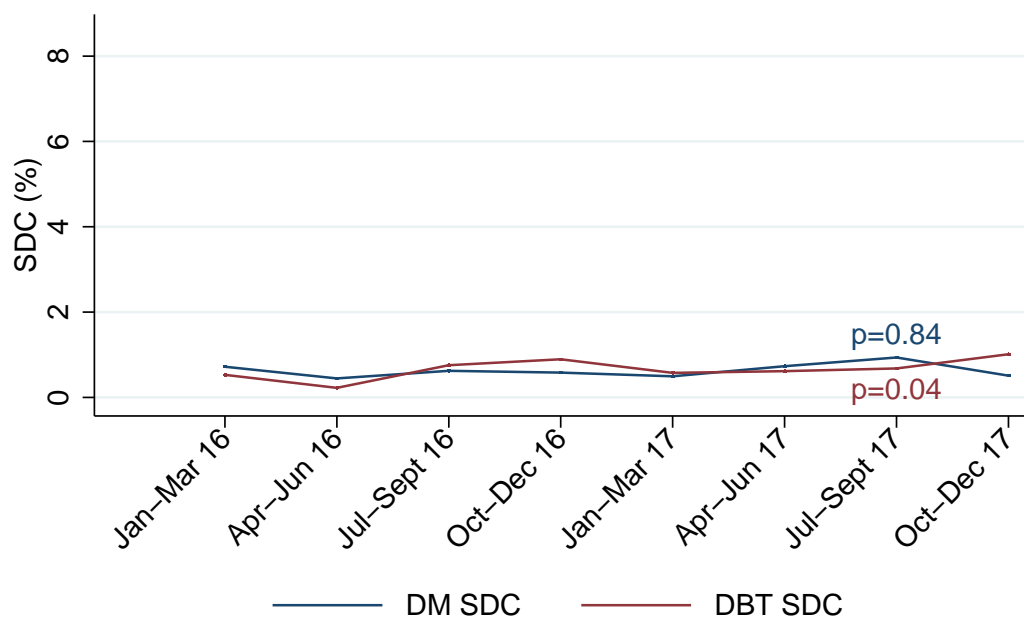
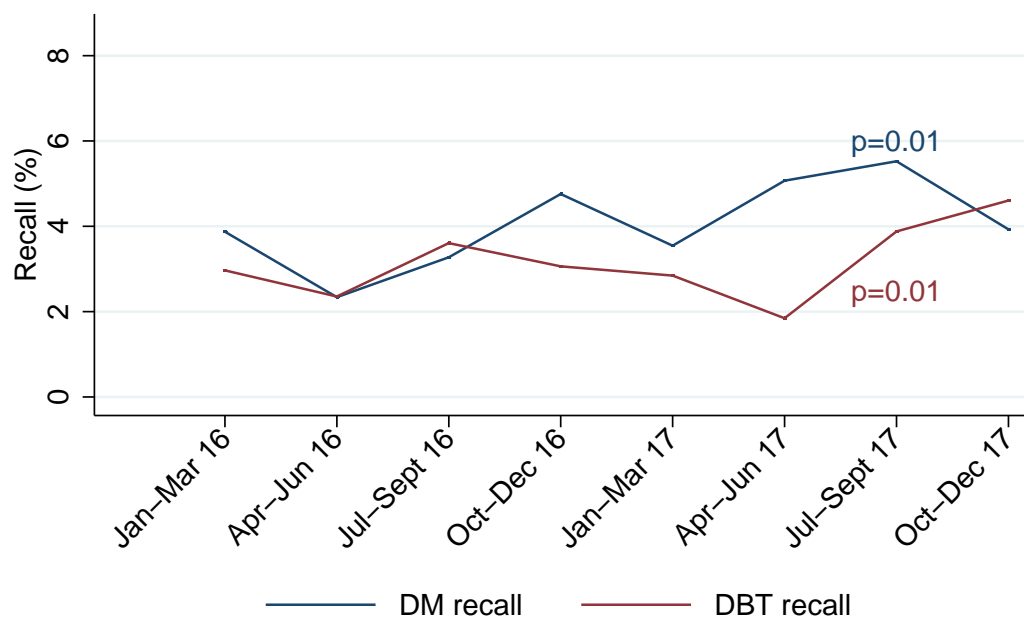
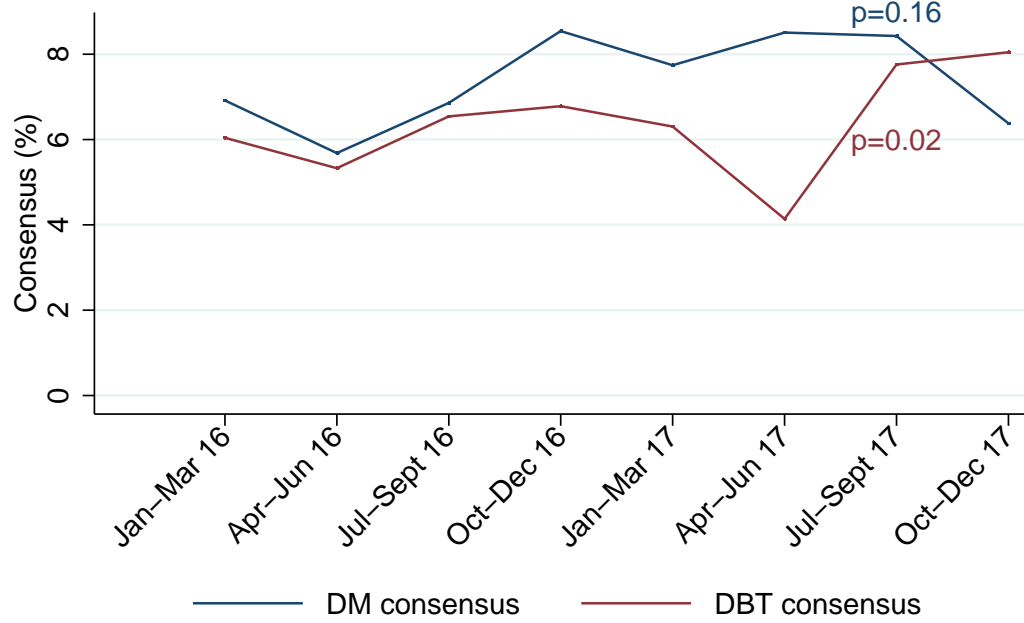
	Total n = 31		DBT n = 15		DM n = 16		p-value
	n	% (95% CI)	n	% (95% CI)	n	% (95% CI)	
Tumor diameter							
Mean, SD (mm)	29	27.6, 25.9	13	26.4, 15.7	16	28.5, 32.5	0.83 [€]
Median, IQR (mm)	29	23.0, 18.0	13	28.0, 10.0	16	18.5, 25.3	
<20 mm	11	37.9 (20.7-57.7)	3	23.1 (5.0-53.8)	8	50.0 (24.7-75.3)	0.14 [‡]
≥20 mm	18	62.1 (42.3-79.3)	10	76.9 (46.2-95.0)	8	50.0 (24.7-75.3)	
Information not available	2		2		0		
Grade							
1	3	10.0 (2.1-26.5)	2	13.3 (1.7-40.5)	1	6.7 (0.2-31.9)	0.82 [‡]
2	10	33.3 (17.3-52.8)	5	33.3 (11.8-61.6)	5	33.3 (11.8-61.6)	
3	17	56.7 (37.4-74.5)	8	53.3 (26.6-78.7)	9	60.0 (32.3-83.7)	
Information not available	1		0		1		

[‡] p-value for Chi-square test between DBT and DM[€] p-value for t-test of mean tumor diameter between DBT and DM

Figure



Figure



Study protocol

2019-02-15

The Tomosynthesis study in Bergen – the To-be trial

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 (NCT02835625)

Principal investigator: Solveig Hofvind, PhD, Cancer Registry of Norway
solveig.hofvindreftregisteret.no

Synopsis

Study title	Digital breast tomosynthesis – the future screening tool for breast cancer?
Study phase	Data will be collected from a two year recruitment period (2016 and 2017) and for two years after the recruitment period, for estimation of interval and breast cancer in consecutive screening round
Background	Digital breast tomosynthesis (DBT) is an advancement of mammography, and has the potential to overcome limitations of standard digital mammography (DM). The use of DBT+DM roughly doubles the radiation dose compared to DM alone. As a result, synthetic 2D mammograms (SM) was developed using raw-data from the DBT acquisition to minimize the radiation burden to women. As of yet, there is insufficient evidence to draw any conclusions about the overall balance of benefits and harms of using DBT+SM in a population-based screening program.
Study aim	To investigate the potential superiority of first generation DBT+SM versus DM in an organized population-based screening program
Study setting	The breast center at Haukeland University hospital, as a part of the national screening program, BreastScreen Norway
Study design	A large-scale, parallel group, superiority RCT
Outcome measures	<p><i>Primary outcome:</i> Screen-detected breast cancer</p> <p><i>Secondary outcomes:</i> Recalls, positive predictive value of recalls and biopsies, prognostic and predictive tumor characteristics, economical aspects</p> <p><i>Other Outcome Measures:</i> Consensus, time spent on screen-reading and consensus, mammographic features, radiation doses and other early performance measures</p> <p><i>After two years of follow up for the individual women:</i> Interval cancer and screen-detected breast cancer among consecutively screened women</p>
Study population	The target group is 45 000 women aged 50-69. We expect 75% attendance rate in the program, and 90% participation rate in the trial, resulting in 30 000 women in total, 15 000 in each arm
Inclusion/exclusion criteria	<p>Inclusion criteria: Women who attend BreastScreen Norway with a complete screening exam and signed an informed consent</p> <p>Exclusion criteria: Women with breast implants were not considered for participation in the trial. Women who have prior history of breast cancer</p>

	or metastatic melanoma, or who report breast symptoms when attending for screening examination will be screened as usual, but excluded post-randomization.
Randomization	Fully concealed, simple randomization and 1:1 allocation ratio. The intervention will not be blinded
Procedures	Screening with two-view DBT+SM or two-view standard DM. Independent double reading of the screening mammograms, by a pool of eight breast radiologists. All cases with a positive score will be 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	In a population with an estimated screen-detected breast cancer rate of 0.60%, we calculated that with 15,000 women in each arm, we could observe an increase in prevalence from about 0.60% with DM to 0.88% with DBT, with 80% power using a two-sided significance threshold of 5%
Statistical analysis	Variables will be described and tested using chi squared tests, t-tests, one way ANOVA and Z tests. The primary outcome will be analyzed with a log-binomial regression model and presented as crude risk ratios with a 95 % confidence interval
Safety considerations	In addition to adhering to the ethical approvals obtained, an interim analysis will be performed after 1 year and published in a peer-reviewed journal to control radiation dose and selected early performance measures
Project management	Consortium: Haukeland University Hospital Cancer Registry of Norway University of Oslo The consortium appoints a Steering Committee. The project group will be led by PI Solveig Hofvind.
Study sponsor	The Norwegian Research Council