Digital Breast Tomosynthesis

- the future screening tool for breast cancer?

Hildegunn Siv Aase

Thesis for the degree of Philosophiae Doctor (PhD)
University of Bergen, Norway
2022

UNIVERSITY OF BERGEN

Digital Breast Tomosynthesis

- the future screening tool for breast cancer?

Hildegunn Siv Aase

Thesis for the degree of Philosophiae Doctor (PhD)
at the University of Bergen

Date of defense: 31.10.2022

Year: 2022

Title: Digital Breast Tomosynthesis

Name: Hildegunn Siv Aase

Print: Skipnes Kommunikasjon / University of Bergen

Scientific environment

Professor Solveig Hofvind and Professor Ingfrid S. Haldorsen supervised me in the work of this PhD-project, performed within Department of Clinical Medicine (K1), University of Bergen. My main scientific environments have consisted of Section of mammographic screening at the Cancer Registry of Norway and Mohn Medical Imaging and Visualization (MMIV), Haukeland University Hospital. MMIV was established in collaboration between the University of Bergen and the Department of Radiology at Haukeland University Hospital through financial support from the Trond Mohn Foundation to promote cross-disciplinary imaging research. The Gynaecologic Cancer Research Group at the Department of Clinical Science, University of Bergen has also been an important part of my scientific environment. The Cancer Registry and MMIV is closely linked to the Breast Centre, Department of Radiology, at Haukeland University Hospital. Professor Hofvind is Head of BreastScreen Norway, and professor Haldorsen is Head of MMIV.

The PhD project was funded by the Research Council of Norway (the To-Be1-trial; project number 24791/H10) and Department of Radiology at Haukeland University Hospital.

UNIVERSITY OF BERGEN

BreastScreen Norway

HELSE BERGEN Haukeland University Hospital

The Research Council of Norway

Cancer Registry of Norway

Acknowledgements

The To-Be-trials were initiated and skillfully drafted by my main supervisor Solveig Hofvind after new mammography machines with options for tomosynthesis, were installed at our screening-site in Bergen in November 2014. Solveig, this thesis would never have been initiated or fulfilled without your everlasting commitment, enthusiasm, support and inspiration. Thank you for giving me the opportunity to become a PhD-student even though I was an established breast radiologist without much experience in research before start-up. In spite of challenges because of the distance between you (working at the Cancer Registry of Norway in Oslo) and me (at the Breast Centre at Haukeland University Hospital in Bergen), you have always supported me. Through visits, invitations, Skype-meetings, some 100- emails, phone-calls and text-messengers, during day-time, evenings, weekends and holydays we have had a lot of contact. This thesis, or the To-Be-trials, would not have happened, and BreastScreen Norway would not have been anything similar to what it is, without you.

I would also like to give special thanks to Ingfrid Salvesen Haldorsen, my enthusiastic and very positive co-supervisor. Ingfrid, your warm and caring personality combined with your highly evaluated professional skills as a researcher and radiologist subspecialized in women cancer imaging has been invaluable. You have wisely guided studies, revised manuscripts, and continuously encouraged me.

I will also like to give special thanks to Aslak Aslaksen, my boss and head of Department of Radiology. Thank you for always supporting me and for giving me the opportunity to finish my PhD, in spite of busy times at the Breast Centre.

Berit Hanestad, without your positive personality and great skills as a radiographer and research assistant in the project the last years would have been much more challenging, and definitely a lot more boring. Thank you for all your help and enthusiasm and for your friendship.

To my co-authors; Åsne Holen, Anders S. Danielsen, Solveig Roth Hoff, Tone Hovda, Nataliia Moshina, Nehmat Houssami, Sofie Sebuegård, Cecilia Sandvik, Kristin Pedersen, Sophia Zackrisson and Chrisoph I. Lee: Thank you for being a part of this adventure and for sharing your experience and knowledge with me.

Thank you to all my colleagues at the Breast Centre in Bergen and at the Cancer Registry for your support and enthusiasm for the study.

I would like to thank my mother, brother and friends for putting things in perspective and making me see what is important in life. Mari, Marit, Kari and Grethe; thank you for keeping me social and active during these years.

I would like to thank my four grown up children Helene, Eivind, Vegard and Jørgen for accepting that your mother as a main rule has been sitting working at her lap-top during the last years. Even though I am very engaged in my work and my research, I think you all know you are the most important part in my life. I love you all.

Finally, Erling, without your acceptance for having a wife working more than nine to five, this thesis would never have been finished. Thank you for making more than your fair share of housework, and for your love, support and trust.

My motivation for the project

In 1994, soon after finishing medical school, I started working at Department of Radiology, Haukeland University Hospital. The Brest Centre at Haukeland University Hospital started mammographic screening in 1996. Thus, from the very early days in my career as a medical doctor, I was introduced to mammography and mammographic screening. The Breast Centre at Haukeland University Hospital has ever since been part of my daily work.

Through my work, and from studies, I have learned that not all breast cancers are detected with mammography. Even more often a finding on mammography turn out to be no pathology at further assessment. As a radiologist I was curious whether the "3D-method", digital breast tomosynthesis, could be a better tool to reveal more aggressive breast cancers in screening compared with digital mammography (2D).

I have also been curious about research, and I have been increasingly engaged in optimization of systems rather than just describing yet another mammogram. Also health economy, and prioritization on what to do next to improve health care, is important. More knowledge and high quality health care systems are keystones for achieving better health care and breast cancer care for women.

The To-Be trials started in 2016. I am grateful for the opportunity to participate in the research and in the practical part of the large scale randomized controlled trial To-Be1 and in the follow up trial, To-Be2. It has been a very exciting journey, including a lot of DBT- and DM-readings together with a steep learning curve in research. My hope is that our trials and our contribution in research through the To-Be-trials will contribute to more knowledge in how to further improve breast cancer screening for women in Norway and in the rest of the world. Through research it is possible to achieve higher efficacy in breast diagnostics, in screening and clinical setting, at acceptable costs benefiting all women.

Abbreviations

DM	Digital Mammography
DBT	Digital Breast Tomosynthesis
SM	Synthetic two dimensional Mammography
PPV	Positive Predictive Value
BI-RADS	Breast Imaging - Reporting and Data System
ACR	American College of Radiology
AI	Artificial Intelligence
MB	Megabyte (=1000 Kilobyte)
To-Be1	Tomosynthesis trial in Bergen (To-Be, 2016-2017)
To-Be2	Tomosynthesis trial in Bergen (2018-2019)
Min:sec	Minutes and seconds
MRI	Magnetic Resonance Imaging
DCIS	Ductal Carcinoma In Situ
MDT	Multi-Disciplinary Meeting
US	United States
mGy	Milligrey
BMI	Body mass index

Abstract in English

Background: Breast cancer is the most common cancer and one of the leading causes of cancer deaths in Norway and globally. Mammographic screening aims for early detection of breast cancer and reduced mortality from the disease. Studies have shown higher rates of screen-detected cancers for digital breast tomosynthesis including ~200-250 images compared to standard digital mammography (DM) including four images. We performed a randomized controlled trial (RCT), the Tomosynthesis trial in Bergen (To-Be1), were the aim was to compare early performance measures for digital breast tomosynthesis including synthesised 2D images (DBT) versus DM in screening. This thesis includes three studies with the following aims:

Study 1: To compare preliminary results of reading time, radiation dose, consensus and recall for DBT and DM after the first year of To-Be1.
Study 2: To compare recall, false positive screening results and screen-detected cancers by automated mammographic density (Volpara density grade, VDG 1-4) and screening technique (DBT versus DM).
Study 3: To investigate distribution of mammographic features in women recalled after screening with DBT versus DM and assess associations between mammographic features and final outcome of the screening examination.

Method: All women who attended the screening unit in Bergen during 2016-2017 as part of BreastScreen Norway (n=32 976) were invited to participate in To-Be1. In total, 89.3% of the women accepted the invitation and were randomized to undergo either DBT or DM. After independent double reading with consensus, results for DBT were compared with DM. Mammographic density were described by VDG 1-4 which are analogue to the categories in the BI-RADS 5th edition. The radiologists classified the mammographic features of recalled women according to a modified BI-RADS scale. We presented descriptive results and used t-tests to test for means, and chi-squared tests for categories with corresponding 95% confidence intervals (CI).

Log-binominal regression models were used to estimate relative risks. A p-value lower than 0.05 was defined as statistically significant. We used STATA software.

Results: Study 1: Mean reading time was 1:11 min:sec for DBT versus 0:41 min:sec for DM in the first year of To-Be1. Mean glandular dose did not differ statistically for women screened with DBT (2.96 mGy) versus DM (2.95 mGy). Recall was 3.0% for DBT and 3.6% for DM in the first year of To-Be1.

Study 2: Recall rate for women with VDG 1 was 2.1% for DBT and 3.3% for DM, while it was 3.2% for DBT and 4.3% for DM for women with VDG 2. The rate of false positive screening results was 1.6% for DBT and 2.8% for DM for women with VDG 1. For women with VDG 2 it was 2.4% for DBT and 3.6% for DM. No statistical difference in screen-detected cancers was observed between DBT and DM in any density categories. Adjusted relative risk of recall, false positives and screen-detected cancers increased with VDG for DBT. No difference was found for DM.

Study 3: The study included 182 screen detected cancers (n=95 DBT and n= 87 DM). 36.8% of those detected with DBT was spiculated mass, while it was 18.4 % for DM. Calcifications was the most frequent feature for breast cancer among those screened with DM (23.0%), which did not differ statistically from the 13.7% for DBT. Asymmetry, indistinct and obscured mass was less frequent in women with a false positive screening result after screening with DBT versus DM.

Conclusion: Results from To-Be1 indicated DBT to be as least as good as DM in terms of recall and cancer detection, which means that DBT is safe for the women. DBT was superior to DM in women with VDG 1 and 2 (lower recall, fewer false positives, no difference in cancer detection). However, time spent on initial screen reading and on consensus was longer for DBT compared with DM. More knowledge of the differences in distribution of mammographic features and their association with screening outcome, might contribute to further improve the benefits of DBT as a screening tool for breast cancer.

Abstract in Norwegian

Bakgrunn: Brystkreft er den vanligste kreftformen blant kvinner og en av de hyppigste årsakene til kreftdødsfall i Norge og globalt. Målsettingen med mammografiscreening er å oppdage brystkreft i et tidlig stadium og redusere dødeligheten av sykdommen. Studier har vist høyere deteksjon av screeningoppdagede krefttilfeller med digital brysttomosyntese som inkluderer ~200-250 bilder sammenlignet med standard digital mammografi (DM) med fire bilder. Vi utførte en randomisert kontrollert studie (RCT), Tomosyntese-studien i Bergen (To-Be1). Målsettingen med studien var å sammenligne tidligindikatorer i screening ved bruk av digital brysttomosyntese i kombinasjon med syntetiske 2D-bilder (DBT) versus standard DM. Avhandlingen inkluderer tre studier med følgende mål:

Studie 1: Å sammenligne lesetid, stråledose, konsensus og tilbakekalling ved bruk av DBT og DM etter det første året av To-Be1.
Studie 2: Å sammenligne tilbakekalling, falske positive screeningsresultater og screeningoppdaget kreft for kvinner med ulik mammografisk tetthet målt automatisk (Volpara tetthetsgrad, VDG 1-4) og med ulike screeningteknikker (DBT versus DM).
Studie 3: Å undersøke fordeling av mammografiske funn hos kvinner tilbakekalt etter screening med DBT versus DM og analysere sammenhenger mellom mammografiske funn og det endelige resultatet av screeningundersøkelsen.

Metode: Alle kvinner som deltok i screening utført i Bergen i løpet av 2016-2017 som en del av Mammografiprogrammet (n=32 976) ble invitert til å delta i To-Be1. Totalt aksepterte 89,3 % av kvinnene invitasjonen og ble randomisert til DBT eller DM. Etter uavhengig dobbelttyding med konsensus ble resultater etter DBT sammenlignet med DM. Mammografisk tetthet ble oppgitt som VDG 1-4, som er analog til kategoriene i BI-RADS´ 5. utgave. Radiologene klassifiserte mammografiske funn hos etterinnkalte kvinner etter en modifisert BI-RADS skala. Vi brukte deskriptive analyser og t-test for å sammenligne gjennomsnittsverdier, samt kji-kvadrat-test med tilhørende 95% konfidensintervall (KI) for å sammenligne kategorier. Log-binominale regresjonsmodeller ble brukt for å estimere relativ risiko.

En p-verdi lavere enn 0,05 ble definert som statistisk signifikant. Vi brukte statistikkprogrammet STATA.

Resultater: Studie 1: Gjennomsnittlig lesetid var 1:11 min:sek for DBT og 0:41 min:sek for DM i det første året av To-Be1. Det var ingen statistiske forskjeller i gjennomsnittlig stråledose for noen av tetthetskategoriene for DBT (2.96 mGy) versus DM (2.95 mGy). Tilbakekallingen var 3,0 % for DBT og 3,6 % for DM etter det første året med To-Be1.

Studie 2: Etterundersøkelsesraten for kvinner med VDG 1 var 2.1% for DBT og 3.3% for DM, mens den var 3.2% for DBT og 4.3% for DM for de med VDG 2. Raten av falske positive screening resultater var 1.6% for DBT og 2.8% for DM for kvinner med VDG 1. For kvinner med VDG 2 var den 2.4% for DBT og 3.6 for DM. Ingen statistiske forskjeller i screeningoppdaget kreft ble funnet mellom DBT og DM for noen av tetthetskategoriene. Justert relativ risiko for tilbakekalling, falskt positivt screeningsresultat og screeningoppdaget kreft økte med VDG i DBT, mens det ikke ble funnet forskjeller i DM.

Studie 3: Studien inkluderte 182 screeningdetekterte krefttilfeller (n=95 for DBT og n=87 for DM). Blant disse var 36,8% spikulerte masser for DBT mens det var 18,4% for DM. Kalk var det hyppigste mammografiske funnet for brystkrefttilfeller for de som var screenet med DM (23%). For DBT var andelen på 13,7%. Asymmetri, uskarp og skjult masse var mindre hyppig hos kvinner med et falsk positiv screening resultat etter screening med DBT versus DM.

Konklusjon: Resultater fra To-Be1 indikerte at DBT var minst like god som DM når det gjelder etterundersøkelser og deteksjon av brystkreft, som betyr at DBT er trygt å bruke i screening. DBT var bedre egnet enn DM for kvinner med VDG 1 og 2 med hensyn til etterundersøkelsesrate og falske positive, mens deteksjon av brystkreft ikke var forskjellig. Det tok lengre tid å lese DBT enn DM bilder, og konsensus tok lengre tid med DBT. Mer kunnskap om forskjeller i mammografiske funn og sammenheng med screeningsresultater for DBT versus DM kan bidra til å ytterligere forbedre fordelene med DBT som et screeningverktøy.

List of Publications

This thesis is based upon three publications referred to in the text by their respective roman numerals:

I. **Aase HS**, Holen ÅS, Pedersen K, Houssami N, Haldorsen IS, Sebuødegård S, Hanestad B, Hofvind S: "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***. 2019; 29:1175-86.

II. Moshina N, **Aase HS,** Danielsen AS, Haldorsen IS, Lee CI, Zachrisson S, Hofvind S: "Comparing screening outcomes for digital breast tomosynthesis and digital mammography by automated breast density in a randomized controlled trial – results from the To-Be trial". ***Radiology***. 2020; 297:522-531

III. **Aase HS**, Danielsen AS, Hoff SR, Holen ÅS, Haldorsen IS, Hovda T, Hanestad B, Sandvik CK, Hofvind S: "Mammographic features and screening outcome in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography". ***European Journal of Radiology***. 10.1016/j.ejrad.2021.109753:109753

Reprints were made with permission from European radiology, Radiology and European Journal of Radiology.

Contents

1. INTRODUCTION

1.1 Breast cancer epidemiology

Breast cancer is the most frequent cancer in women in Norway and worldwide [1]. In Norway 3424 women were diagnosed with breast cancer in 2020, while in 2019, the year before the Covid-pandemic, 3726 women were diagnosed [2; 3]. Breast cancer accounted for about 22% of cancers in Norwegian women during 2016-2020. In Norway ~1/3 of all breast cancers are diagnosed in women participating in BreastScreen Norway [4].

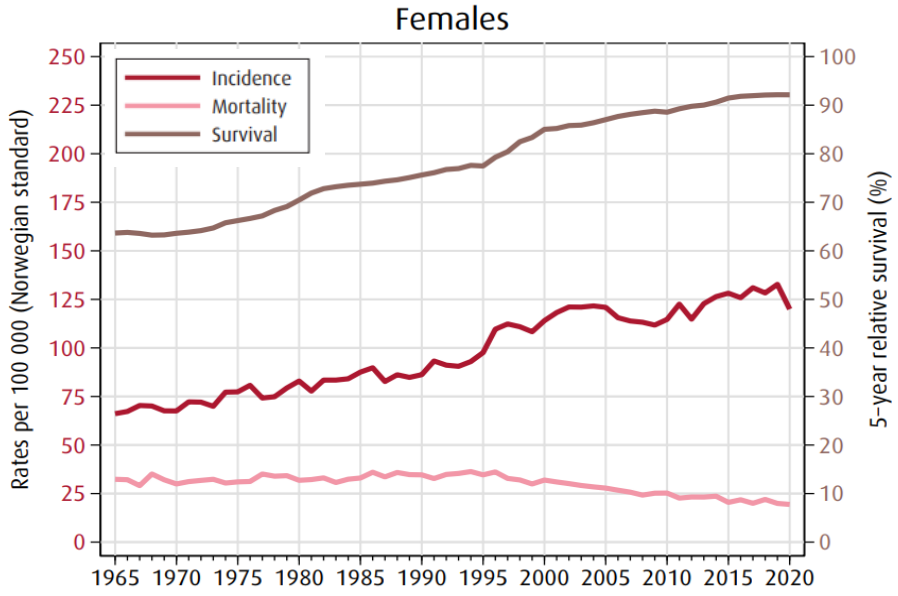

Figure 1: Trends in breast cancer incidence and mortality rates and 5-year relative survival in Norwegian women [3].

The probability of surviving the first five years after being diagnosed with breast cancer relative to a comparable group of women without breast cancer was 92.0% in 2020 [3]. Survival from breast cancer has increased during the last decades (Figure 1). Nevertheless, breast cancer is one of the cancers responsible for most cancer

deaths in women; lung cancer (20%), breast cancer (12%), colorectal (12%) and pancreatic cancer (7%) [3]. Breast cancer mortality has decreased during the last decades (Figure 1). The main reason for the decline is improved treatment and early detection by screening [5-7]. Increased breast awareness and improved diagnostics might have contributed to further improve results.

The prognosis of breast cancer is associated with stage at diagnosis (Figure 2), which is affected by early diagnosis. In Norway during the time period 2016-2020, 5-year relative survival was 100% in women with tumor size less than 2 centimeters and no lymph node involvement (stage 1), versus 34% for those with distant metastases at diagnosis (stage 4) [3].

Screened women have 20-30% lower breast cancer mortality compared to non-screened women [5-8]. During 2020 in total 684 996 breast cancer deaths occurred among women worldwide [9]; 601 in Norway, (591 in women and 10 in men) [3].

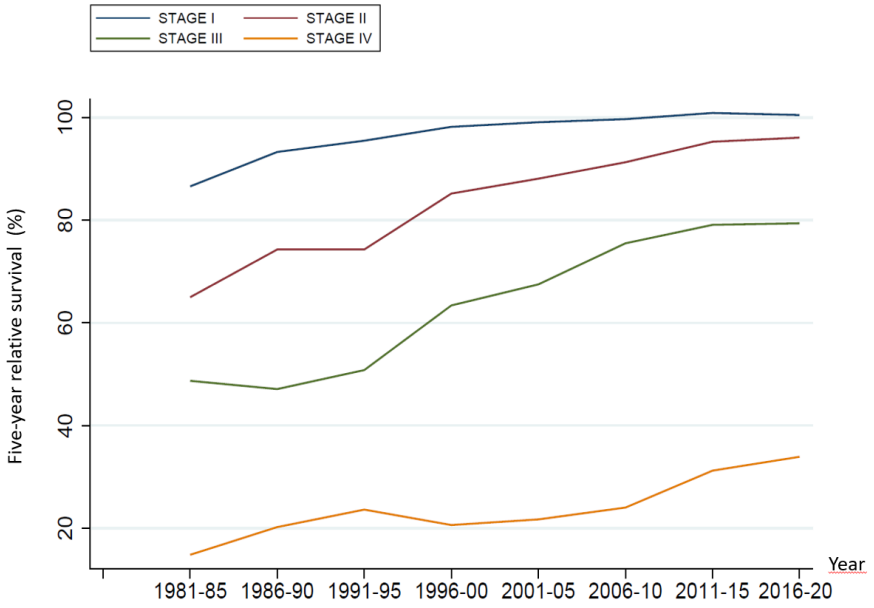

Figure 2: Five-year relative survival for women diagnosed with breast cancer in Norway by stage and diagnosis period [3].

1.1.1 Risk factors

There are several known risk factors for breast cancer. High age is a strong risk factor (Figure 3). Other non-modifiable risk factors are low age at menarche and high age at menopause, high mammographic density and family history of breast cancer [10]. Examples of modifiable factors known to increase breast cancer risk are use of hormonal replacement therapy, nulliparity, high age at first birth and high body mass index (BMI), while physical activity and more child births and breast feeding reduces risk [10-17]. Furthermore, previous assessment after suspicious findings on the screening mammogram without diagnosing breast cancer (false positive screening) and benign breast biopsy have been shown to be risk factors for breast cancer [18-20].

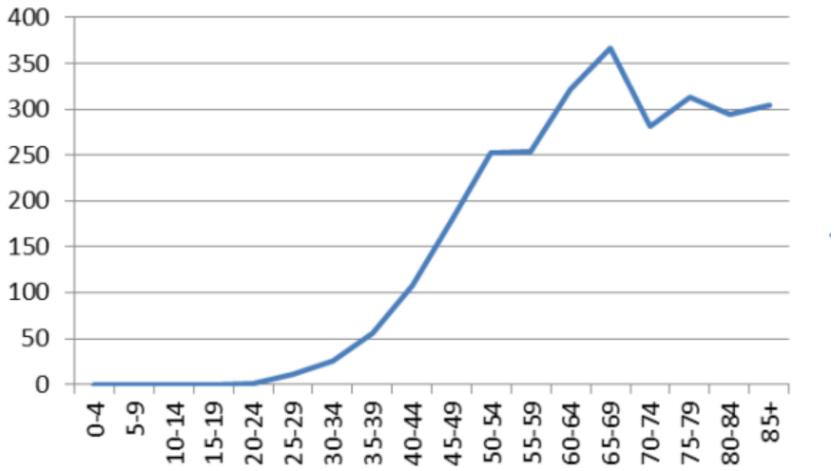

Figure 3: Age-specific incidence rate per 100 000 person years in Norway, 2014-2018 [21].

1.2 Diagnosing breast cancer

1.2.1 Clinical examination, imaging and biopsy (triple test)

Breast cancer is a heterogeneous disease in terms of clinical manifestation, tumor biology and visibility on mammography and on other imaging modalities. Therefore, triple assessment test, including clinical examination, radiological imaging and histopathology, is considered standard diagnostic procedure for women referred for

assessment of palpable breast lumps or other breast symptoms. A clinical assessment evaluates the texture, the size and position of the lump, and can give some information on the likelihood of malignancy [21]. Diagnostic imaging, traditionally being mammography and ultrasound, is another part of the assessment. In recent years additional imaging by contrast enhanced mammography or magnetic resonance imaging (MRI) has been increasingly used. If clinical examination and/or radiology confirms a pathologic finding, a core needle biopsy or a vacuum biopsy is performed to confirm the histological diagnosis [21; 22]. After imaging and biopsy are performed, the multidisciplinary team (MDT), consisting of radiologists, breast surgeons, pathologists and oncologists decide on optimal treatment if breast cancer is confirmed [21; 22].

1.2.2 Digital mammography (DM)

Two-view screen film mammography was introduced in the 1960s as an imaging method for diagnosing symptomatic breast cancer [23]. In the 1970s and 1980s the method was increasingly used as a screening tool for early diagnosis of the disease [23]. In the period 2000-2015 screen film mammography was replaced by digital mammography (DM) [24; 25], which is currently the standard screening method in most European countries. Two-view DM in craniocaudal- and mediolateral oblique view, in total four images, is standard examination in screening and often the first part of the imaging examinations performed in symptomatic women. In DM the x-ray tube is always positioned 90 degrees on the detector, with craniocaudal and mediolateral oblique x-ray beam usually at an angle of 60 degrees for standard imaging [26].

For all 2D-imaging there is a limitation regarding the superimposition of tissue that occurs when a three-dimensional breast is presented on a two-dimensional detector, being most limiting for mammographically dense breasts [15; 27; 28]. Thus, even though there are no suspicious findings on the DM there is still a possibility of breast cancer, especially in women with mammographic dense breast.

1.2.3 Digital breast tomosynthesis (DBT)

Digital breast tomosynthesis is a rather new quasi three-dimensional imaging technique where the x-ray tube moves in an arch above the breast while taking several low-dose images which is later reconstructed into several thin planes through the breast [29; 30], (Figure 4). Usually, the thin planes are read in combination with a two-dimensional view, like DM, however this doubles radiation dose. Synthetic two dimensional mammogram (SM), have recently been developed, and is aimed at replacing the additional DM [31], using no more radiation, since SM is reconstructed from the same set of raw data as the DBT-planes. For a two-view digital breast tomosynthesis included SM (hereafter DBT) of both breasts, the total number of images is ~200-250 images; a DM-view is replaced with ~50 (45-70) thin planes (each often 1 mm), a SM-view, and eventually also some 10 mm slabs.

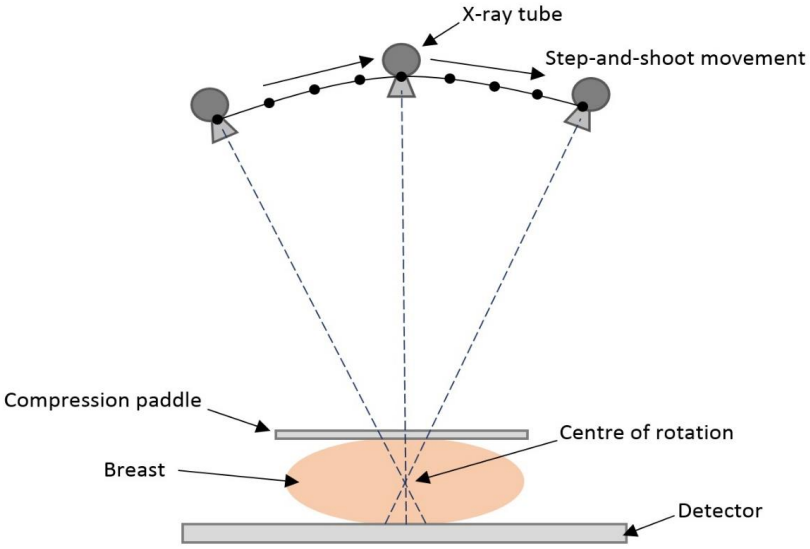

Figure 4: Principle of DBT-uptake; step and shoot movement with 9 low-dose images acquired, as used in the GE-system. Other systems may use continuous movement and a different number of low-dose images [32].

DBT has the potential to reduce the effect of superimposed tissue which is known to be a challenge in DM [33]. DBT has been introduced at several breast centres in Norway as well as in Europe and the United States during the last decade, as a

supplement or instead of DM for symptomatic mammography and for women recalled due to suspicious screening mammograms.

Data storage needed for a DBT varies between vendors [34]. A DBT data set requires more storage capacity in the hospitals PACS systems compared with DM: if stored in a full resolution 1000-3000 MB can be needed for a two-view DBT of both breast compared to 32-96 MB for a DM exam. The size of the images depends on pixel-size, breast area in contact with detector, the compressed thickness of the breast and slice-thickness [34; 35].

Several vendors offer DBT, however, even though the term is similar, vendors use different techniques regarding sweep angle of the x-ray tube, numbers of low-dose projections acquired, how the x-ray-tube moves, and other factors (Table 1) [34; 36]. Worth noticing, the quality of SM and the thin plane images may also differ between DBT-systems due to different techniques in DBT uptake, reconstruction and postprocessing.

The DBT radiation dose has been estimated to be typically ~2.3mGy per view in an average-sized breast, which is 1-1.5 times the dose of a DM-image [33]. However, the radiation dose needed for DBT differ between vendors [37]. The DBT-systems available are set up differently regarding what dose to determine from the automatic exposure control (AEC) software; therefore dose is difficult to compare across systems.

Thus, it is important to acknowledge the differences between vendors, and to use caution when extrapolating results from one DBT-system to another.

Table 1: System specifications for some FDA-approved DBT-Vendors, updated December 2020 [32; 34].

Manufacturer	**Hologic**	**Hologic**	**GE**	**GE**	**Siemens**	**Fujifilm**
Model	Selena	3Dimensions	SenoClare	Senographe Pristina	Mammomat	ASPIRE Cristalle
Year FDA approved	2011	2018	2014	2017	2017	2017
X-ray tube						
Target/filter	W/AI	W/AI	Mo/Mo Mo/Rh Rh/Rh	Mo/Mo Rh/Ag	W/Rh	W/Al
Detector						
DM pixel size (microns)	70	70	100	100	85	50
Pixel binning for DBT	2x2	None	None	None	None	2x2 (ST) None (HR)
Effective DBT resolution	100	70	100	100	85	150 (ST) 100(HR)
Acquisition technique						
Sweep angle (*)	15	15	25	25	50	40 (HR) 15 (ST)
X-ray tube motion	Continous	Continous	Step-and-shoot	Step-and-shoot	Continous	Continous
Number of projections	15	15	9	9	25	15
DBT-only scan time (sec)	3.7	3.7	10	10	29	9(HR) 4(ST)
Image reconstruction						
Slize thickness	1 mm	1 mm	0.5-1 mm	0.5-1 mm	0.2*	1 mm
FDA-approved synthetic 2D	C-view	Intelligent 2D	V-preview	V-preview	Insight 2D	S-view
DBT-guided biopsy	Yes	Yes	No	Yes	Yes	Yes
Mean glandular dose for a 45 mm equivalent breas thickness (actiual clinical values may differ)	1.37 mGy (as per NHS BSP evaluation)	No data available	1.09 mGy (as per NHS BSP evaluation)		1.67 mGy (as per NHS BSP evaluation)	1.23-1.88 mGy depending on mode used (dose from Fudjifilm)
Storage space required compared to 2D	20-30x depending on compression formats	No data available	~26x dependent on user/archive storage options		10 x	50x

Abbreviations: Ag, silver; Al, aluminium; FDA, Food and drug Administration, HR, High-resolution mode; Mo, molybdenum; N/A, not applicable; Rh, rhodium; ST, Standard mode; W, tungsten. T, standard mode; HR, high definition mode; NHS BSP, National Health Service Breast Screening Programme.
* merged into 2.0 mm slices for viewing.

1.2.4 Supplementary imaging

Additional views and imaging techniques are used for diagnostics, depending on symptom location and the mammographic feature detected. Cleopatra-view is used if there is need for better visualisation in the lateral part of the breast. Mediolateral image and cone might be used for better visualisation of masses, calcifications and asymmetries. Magnification, with or without cone, may better characterize calcifications. Ultrasound has become the second most important imaging method in breast diagnostics during the last decades. Increasing number of patients also undergo magnetic resonance imaging (MRI) or contrast enhanced mammography to better visualize extent of the pathology and tumor boundaries.

1.2.5 Histopathology and staging of breast cancer

Treatment is recommended for all women diagnosed with invasive breast cancers as well as for ductal carcinoma in situ (DCIS) and some special types of lobular carcinoma in situ [21; 22]. The clinical tumor-node-metastasis (cTNM) classification system, incorporating information from triple assessment test and prognostic and predictive tumor characteristics are used to guide the MDT on what treatment to recommend [21; 22; 38] (Table 2). The TNM-classification system defines the stages of the breast cancer based on a) description of size of the tumor (T), b) involvement of regional lymph nodes (N), and c) the absence or presence of distant metastases (M) [38]. Prognostic tumor characteristics (Table 2) provides information about the prognosis of the patient, while predictive tumor characteristics (Table 2) indicates how the cancer will respond to treatment [22].

Table 2: Prognostic and predictive tumor characteristics of invasive breast cancers [22].

	Prognostic	Predictive
Tumor diameter	+	-
Grade of invasive tumor	+	-
Estogen receptor (ER)-status	+	+
Progesterone receptor (PR)-status	-	+
HER2-status	+	+
Ki-67 proliferation	+	-
Subtype	+	+
Lymph node involvement	+	-

To define histological grade of a tumor the Nottingham histologic scoring system is used. Grade 1 cancers are in general less aggressive, and more often estrogen positive compared with grade 2 and 3. Grade 3-cancers are more aggressive, more often "triple negative" (negative for estrogen and progesterone receptor and negative HER2-status) and have a higher risk of relapse after treatment compared with grade 1 cancers [38]. Ductal carcinoma in situ (DCIS) is classified as grade 1-3 according to van Nuys classification system [39].

Further, the classification system for immunohistochemical subtypes proposed after St. Gallen International Expert Consensus, using surrogate molecular subtyping from routine immunohistochemical- and in situ hybridisation analyses, are increasingly used to guide optimization of treatment (Table 3)[40; 41]. The majority of breast cancers are estrogen positive, and luminal A-like cancer, which is the most frequent subgroup [42-44]. The rather few cancer cases in the remaining subgroups may cause issues regarding small numbers in statistics and therefor can be collapsed in a "non-luminal A-like" group. HER2 positive and especially triple negative cancers have been shown to be associated with a poor prognosis, while luminal A-like cancers have a more favourable prognosis [42; 44].

Table 3: Subtypes after St Gallen [41].

	ER	PR	HER2	Ki67	
Luminal A	+	+	-	low	Luminal A-like
Luminal B HER2-	+	any*	-	any*	Non-luminal A-like
Luminal B HER2+	+	any	+	any	
HER2+	-	-	+		
Triple negative	-	-	-		

*at least PR- or high Ki67

1.3 Screening for breast cancer

Due to limited knowledge on the causes of breast cancer on an individual level, there are limited opportunities regarding primary prevention for breast cancer. Therefore,

secondary prevention is performed. Screening for breast cancer has been introduced in most European countries as well in the US and other parts of the world [45]. The goal is to detect the cancer at an early, asymptomatic stage, and thereby reduce mortality. Further, women with early detected tumors are expected to have less aggressive treatment, which might reduce morbidity from the disease, and reduce the immediate and late side effects of the treatment [7; 45].

According to the World Health organization`s criteria for screening, the screening test should be used in a population where breast cancer is an important health problem (Table 4). The test should be a suitable test which is effective, feasible to perform and acceptable for the population (Table 4) [46].

Table 4: The World Health organization`s 10 principles of early disease detection (Wilson-Junger criteria) [46], and the 6 additional Norwegian criteria for screening [47].

1	The condition sought should be an important health problem
2	There should be an accepted treatment for patients with the recognised disease
3	Facilities for diagnosis and treatment should be available
4	There should be a recognisable latent or early syptomatic stage
5	There should be a suitable test or examination
6	The test should be acceptable to the population
7	The natural history of the conditioan, including development from latent to declared disease should be adequately understood
8	There should be an agreed policy on whom to treat as patients
9	The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
10	Case-finding should be a continuous process and not a "once and for all" project
Norwegian criteria added in 2014	
11	The health benefits should outweigh the harms
12	The protection of personal privacy and adherence to the law should be ensured
13	The progrem should be ethically acceptable
14	Information about partisipation should be evidence-based and facilitate an informed choice about participation
15	The program should be cost effective
16	There should be a plan for programme administration, quality assurance and evaluation

Factors affecting specificity, sensitivity, positive- and negative predictive value (Table 5) as well as the rate of women with a false positive screening result are thus important to consider. A false positive screening result is defined as a women recalled

from screening without being diagnosed with breast cancer at assessment or during a 6 month period after screening. Feasibility and cost-effectiveness (Table 4; number 9, 11 and 15) are also important when evaluating a screening program.

Table 5: Overview over the terms true positive, false positive, false negative and true negative used in a screening test for breast cancer. Below the definitions of sensitivity and specificity is presented mathematically.

	Breast cancer	No breast cancer
Positive screeningtest (recalled)	True positive (TP)	False positive (FP)
Negative screening test (not recalled)	False negative (FN)	True negative (TN)

$$\text{Specificity} = \frac{\text{True negative (n)}}{\text{True negative (n)} + \text{False positive (n)}}$$

$$\text{Sensitivity} = \frac{\text{True positive (n)}}{\text{True positive (n)} + \text{False negative (n)}}$$

$$\text{Positive predictive value (PPV)} = \frac{\text{True positive (n)}}{\text{True positive (n)} + \text{False positive (n)}}$$

$$\text{Negative predictive value (NPV)} = \frac{\text{True negative (n)}}{\text{True negative (n)} + \text{False negative (n)}}$$

As breast cancer is most common in women older than 45-50 years (Figure 3), screening is recommended only for specific age groups. In Norway and several other European countries women aged 50-69 are offered biennial screening with two-view digital mammography in line with the European guidelines [24]. It is, however, increasingly discussed which age-groups that are suitable for screening. If women are invited to screening before their menopause, mammographic density is expected to be higher because of higher levels of circulating estrogen and progesterone. Despite known challenges with higher mammographic density in premenopausal women, both the EuroScreen working group in EU and International Agency for

Research on Cancer (IARC) working group of WHO recommend considering screening also in age groups below 50 years [48; 49]. These expert groups are also positive to screening in age groups above 69 years. Several countries already have included women from age 45 or 40 years (Austria, Czech Republic, Greece, Hungary, Iceland, Portugal, Russian Federation, San Marino, Slovakia, Spain, Sweden) and/or up to 74 or 75 years in the screening program (France, Monaco, San Marino, Sweden (some regions), Switzerland, the Netherlands, United Kingdom) [50].

1.3.1 Screening outcome of DBT versus DM

The aim of screening is to reduce breast cancer mortality. However, screening programmes have to run for more than 10 years before any impact on mortality can be evaluated [51; 52]. Thus, several early performance parameters are of interest when evaluating the programmes, such as breast cancer detection rate, interval cancer rate (cancers diagnosed between screening rounds), rate of overdiagnosis, time spent during screening-uptake, screen-reading and consensus, radiation dose, number of women recalled, the rate of recalled women diagnosed with breast cancer, and the rate of recalled women with a false positive screening result. The early performance parameters are also used in the discussions on which technique is better suitable for breast cancer screening, DM or DBT. They influence sensitivity, specificity and/or costs.

Screening traditions and study-designs

Several studies have compared early performance measures after screening with DBT including DM or SM, versus DM alone [53; 54]: However, study-designs and study populations have differed, hampering comparison and conclusions. Furthermore, screening is differently performed in different parts of the world: In the US, screening is not population based, includes first and foremost single reading, is performed annually, and is characterized by a relatively high recall rate (11%) compared to Europe (3.5%) [54]. In Europe, organized population-based screening is traditionally performed with double reading, a longer (predominantly biennial) screening-interval, and with a lower recall rate than in the US [53; 54]. Furthermore, most DBT-studies

performed in the US are retrospective, while several prospective trials have been performed in Europe [53-55].

A prospective longitudinal cohort study follows a group of individuals (cohort) who differ in respect to certain factors under study, for instance DBT versus DM. The aim is usually to determine how these factors affect certain outcomes, i.e. early performance measures. This method is sometimes used to test a new intervention, as DBT.

In comparison, in a retrospective cohort study groups of individuals who is alike in many ways, but differ by a certain characteristic (e.g. exposed to DBT versus DM), are compared for an outcome of interest (recall rate, breast cancer, interval cancer rate) [56] .

Prospective cohort studies rank above retrospective cohort studies in the evidence pyramid of epidemiologists, with randomized controlled trials (RCT) at the top of the pyramid [56].

In an RCT, subjects in a population are randomly allocated into a study group or a control group which receive or do not receive the intervention. Results are assessed by comparison of rates of outcome for the two groups. The risk of selection biases and confounding is considered to be low.

Screen-detected and interval breast cancer

The heterogeneity of breast cancers is important to keep in mind when assessing the quality of a screening program. Screening aims to detect tumors at an early stage. Especially the aggressive, rapid growing tumors (e.g. invasive, histological grade 2-3, high Ki67, triple negative or HER2+) are important to detect early to reduce breast cancer mortality. Further, women with high mammographic density have an increased risk of a false negative screening result followed by interval- and consecutive round screen detected cancers with large tumor diameter [57; 58]. Thus, it is not only the *number* of screen-detected cancers, but also histopathological

characteristics and tumor diameter at diagnosis that should be considered when evaluating screening programs (Table 2 and Table 3) [40; 41].

Interval cancers has been found to have less favorable tumor characteristics compared with screen-detected breast cancers; larger tumor size, higher proportion of lymph node positive tumors and lower proportion of luminal A-like tumors [19; 59; 60]. Introducing a better screening-tool should putatively decrease the interval cancer-rate compared with the screening method today (DM). Most of the slow growing cancers should be detected in next screening round, and primarily rapidly growing cancers that truly arise between screening rounds should be diagnosed as interval cancers. In BreastScreen Norway the interval cancer rate has been rather stable comprising ~24 % of all the breast cancers diagnosed in the program in the period 1996-2016 [61].

Sensitivity of DBT in combination with DM has been claimed to be superior compared with DM alone in prospective screening trials [62-66]. However, until recently very limited results regarding performance and comparisons between screening performed with DBT-planes including SM, versus DM have been available [31]. The meta-analyses of Marinovich et al and Alabousi et al which both included European as well as US-studies, found increased cancer detection rates after screening with DBT alone or in combination with DM or SM compared with DM [53; 54]. A study-level meta-analysis published in 2021 including prospective European population-based screening-studies with predominantly biennial screening-interval comparing DBT alone or in combination with SM or DM versus DM, and where also information about interval cancer rate was available, reported robust evidence that screening with DBT increased cancer detection rate. However, no apparent effect on interval cancer rate was found in this meta-analysis [67]. Recently the increased detection rate in DBT versus DM has been confirmed in review studies and in a multicentre RCT [55; 68; 69]. The paired studies could not report on differences in interval cancer rate after DBT versus DM because of study design; all women had DBT and DM. Further, no significant reduction in interval cancer rate was demonstrated in the first DBT-studies reporting on interval cancers [70; 71]. However, the interval cancer rate has recently been reported to be lower for DBT-

screened compared with the control group screened with DM in the Malmö trial [72]. Worth noting, the interval cancer rate among the DM-screened was higher compared with other studies. Furthermore, also women in age group 40-49 were invited to screening, only one view were used for DBT, and a different vendor was used than in the other trials [72]. To reveal significant differences in interval cancer rate is challenging, because of the infrequent occurrence.

More studies are needed, preferably randomized trials, using individual personalized (and not study level-) data, studying screen-detected- and interval cancer rates and long-term outcome, such as next round screen-detection cancer rate and next round interval cancer rate.

Overdiagnosis and overtreatment

Overdiagnosis is defined as cancers diagnosed that would never have presented symptomatically during women's lifetime in absence of screening, because of indolent or slow growing tumors [73]. Small dormant tumors consisting of malignant cells with non-aggressive histological characteristics at diagnosis have a higher susceptibility to be overdiagnosed cancers compared with tumors with more aggressive characteristics. The level of overdiagnosis is important when assessing the different screening techniques. However, overdiagnosis is difficult to measure; treatment is recommended in all women with cancer, and whether the slow growing and indolent cancers would not have needed this treatment remains uncertain. When a tumor with non-aggressive histological characteristics is diagnosed it is essential that less aggressive treatment is considered, in order to reduce the risk of immediate and late side effects from treatment [7; 45]. To give a woman an "unnecessary" although correct cancer-diagnosis is undesirable. Studies have suggested different levels of overdiagnosis in screening performed with DM, but in general overdiagnosis has been considered rather low and at an acceptable rate [74; 75]. The increased rates of detected breast cancers in DBT screening trials, without a simultaneous decrease in interval cancer rates, raises the question whether screening with DBT mainly increases the level of overdiagnosis [51; 99; 117]. According to the authors of a meta-analysis of prospective DBT studies, this could suggest, but does not prove, some

level of overdiagnosis in the included DBT screening studies [67]. There is need of more knowledge regarding whether the thin planes in a DBT will result in earlier detection of all types of breast cancers, or mainly the slow growing cancers.

Reading-time and time spent in consensus (arbitration)

One of the arguments for using DM instead of DBT in screening is the reading time. It is important to monitor reading time in terms of administrative planning, economic costs, workload for radiologists and because of limited accessibility of breast radiologists (Table 4; number 9 and 15). Several studies have shown DBT to require longer reading times than DM; the Oslo tomosynthesis screening trial reported 41 seconds more [66], the Malmö trial (only one view DBT) 30 seconds more [64], the STORM trial in Italy 44 seconds more [63] and Dang et al 54 seconds more [76] when reading DBT in a study from the US. Results will be affected by number of views, hanging protocol, power and speed of the IT systems as well as by the reading speed of participating radiologists [63; 64; 66; 76].

In screening programs using double reading with consensus, also the time spent in consensus should be evaluated, as this could be a time-consuming process for at least two radiologists. At the consensus meeting two or more radiologists participate and make the final decision whether to recall women with a positive score in the independent double reading. As far as we are aware, no European screening trials have reported time spent on consensus (arbitration) after screening with DBT. Further, time spent for radiographers during the acquisition of the screening examination should also be considered when evaluating the two techniques.

Radiation dose

The first prospective DBT-studies were performed with DBT-planes in addition to DM for all women [62; 77]. The SM has been established to replace the DM image, and radiation dose seems less of a challenge [31], however, radiation dose still has to be considered. Even though the dose in mammography is low, efforts should be made to reduce ionizing radiation to the minimum necessary to detect cancers, especially in a population based screening programme where most of participants are negative for

cancer. Thus, if mean glandular dose (MGD) is shown to be lower in screening with DM versus DBT, dose can be an argument for continuing with DM in screening, because of the risk of radiation induced breast cancer.

For most, but not all vendors there is a slight dose increase when performing DBT compared with DM [37]. Thus, there are differences between vendors regarding radiation dose needed in DBT in an average breast (Table 1) [34; 37; 78]. Further, the incremental dose needed in DBT compared with DM might vary with compressed breast thickness and mammographic density [37; 79]. Mean glandular dose for a 45 mm equivalent breast thickness has been reported to vary between 1.09 to 1.88 milligrey (mGy) for a DBT view (Table 1) [34].

There is limited evidence regarding the differences in radiation doses in DBT compared with DM in women with different breast densities and performed with equipment from different vendors, this should be explored in more detail [79; 80].

Recall rate and concensus rate

Recall-rates for DBT versus DM are shown to vary within and between screening programs [24; 54]. Different study designs make it challenging to compare such results for studies performed in different screening programs or in different countries. In general, similar or slightly higher recall rates are shown for screening with DBT versus DM in several European studies, while lower rates are reported from the US [54]. When considering the rather lower baseline recall rate for DM in Europe and the traditionally higher rate for DM in the US, this might seem reasonable. Most studies still show higher recall rates for DBT in US compared to Europe [53; 54]. Being recalled for further assessment after screening can be very stressful for women, further, it is resource consuming to do the assessment [81; 82]. The rate of false positive screening result and PPV of recalls should thus be considered when evaluating DBT versus DM in a screening setting. Further, more knowledge are needed regarding recall and consensus rate and the consequences of these on detection rate and interval cancer rate in DBT versus DM.

Storage, IT and technical -issues

A four-view DBT require more storage capacity compared with DM (1000-3000 MB versus 32-96 MB if stored without reversible (lossless) compression) [34; 35]. In population-based screening this is an important issue to consider regarding the feasibility.

Studies and screening programmes investigating if IT-systems could manage high volume screening workload when screening are performed with DBT versus DM are warranted.

1.4 Mammographic density

Mammographic density reflects the proportions of fibroglandular and fatty tissue in the breasts [83]. Fibroglandular tissue is visualized as white (dense) areas on the mammogram, whereas fatty tissue is radiolucent. Mammographic density varies during a woman's life; density increases with breast feeding, while it usually decreases after menopause when fatty tissue replaces fibroglandular tissue [11; 84-87].

Studies have shown 4-6 times higher risk of breast cancer in women with mammographic dense versus fatty breasts [12; 15; 57; 87-90]. Several risk factors have been shown to be associated with mammographic density (see Chapter 1.1.1) [11; 91]. Mammographic density has also been reported to predict breast cancer risk in several studies [12; 14; 15]. Late age at first birth is associated with high mammographic density, while parity and increased number of births are associated with low mammographic density [13; 85]. BMI is inversely associated with mammographic density [92], while postmenopausal hormone therapy with combined estrogen-progesterone increases mammographic density and breast cancer risk [93; 94].

1.4.1 Mammographic density in BI-RADS (Breast Imaging Reporting and Data System)

BI-RADS is an acronym for Breast Imaging - Reporting and Data System, which is a quality assurance tool designed for mammography published by the American College of Radiology (ACR) [95]. Mammographic density is reported in the second step of the standard reporting after BI-RADS which includes: 1. Indication for the mammogram, 2. Breast composition (mammographic density), 3. Important findings (mammographic features), 4. Comparison to previous studies, 5. Final Assessment Category, 6. Give management recommendations, 7. Communicate unsuspected findings with the referring clinician.

BI-RADS is used in the US and in some European countries. In this thesis mammographic density (*2. Breast composition)* and mammographic features (*3. Important findings)* in BI-RADS will be discussed further.

In BI-RADS 5.th edition from 2013 there are four mammographic breast density (composition) categories:

a) The breast are almost entirely fatty. Mammography is highly sensitive in this setting.
b) There are scattered areas of fibroglandular density. The term density describes the degree of x-ray attenuation of breast tissue but not discrete mammographic findings.
c) The breasts are heterogeneously dense, which may obscure small masses. Some areas in the breasts are sufficiently dense to obscure small masses.
d) The breasts are extremely dense, which lowers the sensitivity of mammography.

In the BI-RADS edition from 2003, used in several DBT screening trials [66; 96], the categories were defined according to the overall density resulting in ACR category I–IV; ACR category I: <25% fibroglandular tissue, ACR category II: 25-50% fibroglandular tissue, ACR category III 50-75% fibroglandular tissue and ACR category IV >75% fibroglandular tissue.

1.4.2 Measuring mammographic density

Visual assessment or automated reporting of mammographic density

In the US mammographic density traditionally has been reported by radiologists using BI-RADS. In Norway a 3-point scale reported by the radiologists was earlier used for women recalled for further assessment in BreastScreen Norway (1995-2012). The assessment was changed to the 4-point scale ("a-d"), identical to the scale in BI-RADS 5th edition in 2013 [97]. In general, inter- and intra-reader variability for density assessment has been shown to be substantial [98]. Automated software for assessment of mammographic density has been developed [99-101] and provides higher reproducibility compared with subjective assessments [102; 103]. In the US, breast density legislation has been enacted in 38 states; the mammography-reports presented to women have to include information on risks related to breast density. Thus, the need for a more reproducible automated software system seems obvious [86].

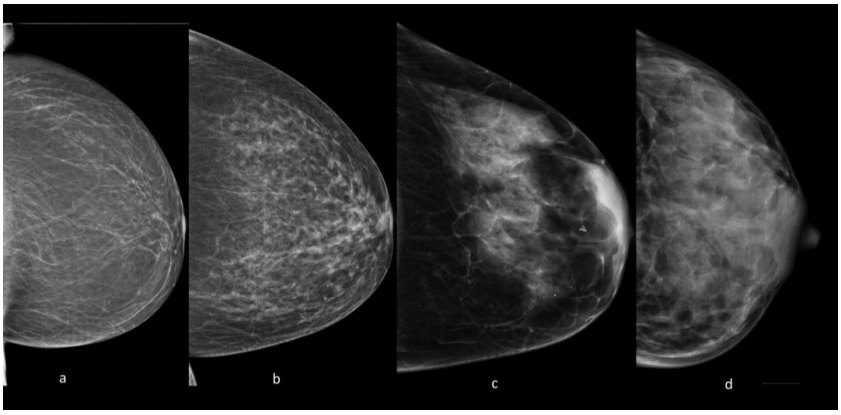

Figure 5: Mammograms in women with BIRADS-densities a, b, c, and d.

There are several automated breast density systems available. Volpara, installed in some breast centres in Norway, is an example of a fully automated software for breast density assessment [104]. Continuous measures of compressed breast thickness, breast volume, fibroglandular volume and volumetric breast density (VBD, percentage of the breast volume) are provided by the software. Volpara uses

information from each pixel regarding density and information on compressed breast thickness to determine the amount of fibroglandular tissue in each pixel. Total amount of fibroglandular tissue in the breast is divided by the whole volume of the breast, resulting in a volumetric breast density (%). Using information from the DICOM-header and the volumetric breast density (%) presented by Volpara, an automated density grade is calculated; Volpara density grade 1-4 which is analogue to BI-RADS 5TH edition a-d [105].

1.4.3 Mammographic density and screening with DM

The visibility of a breast cancer on DM is strongly affected by breast density. Sensitivity of DM is lower in women with dense breasts, because overlapping breast tissue might mask the tumor [106], thus, the challenge of normal breast tissue masking a malignant tumor is considerable [86; 107]. In a fatty breast most features are visible, however, overlapping normal breast tissue might appear as a tumor [33].

As mean age at menopause has been shown to be 51 years [108], most women in age group 50-69 theoretically should have lower breast density compared with women in the premenopausal stage. Still, a substantial rate (28%) of women participating in BreastScreen Norway had breast density in one of the two densest categories [86].

High mammographic density has been shown to be associated with large tumor diameter, lymph node involvement and other poor prognostic and predictive tumor characteristics in women screened with DM [109; 110]. Further, increased risk of recall, biopsy and higher odds of cancer and interval cancer have been found in women with dense breasts compared with women with non-dense breasts [12; 15; 57; 58; 86-90].

The sensitivity of screening with DM is 86-89% for almost entirely fatty breasts [27] but shown to be as low as 50-60% for women with extremely dense breasts [28].

1.4.4 Mammographic density and screening with DBT

Whether DBT is a better technique compared with DM in women with dense breasts has not specifically been studied until recent years. Theoretically, the thin planes in a

DBT should decrease superimposition of normal breast tissue and present tumor-margins better (Figure 6). Thus, DBT may be better suited to discriminate malignant from benign lesions. DBT might have the potential to improve sensitivity and specificity by reducing the overlapping effect of breast tissue occurring in DM, especially in women with dense breasts [33].

Mammographic density has been reported in several prospective DBT screening trials. The Oslo trial reported all four different BI-RADS categories regarding density (edition 2003); fatty, scattered, heterogeneous and extreme, which is similar, but not completely identical to the categories in the 5th edition of BI-RADS. The Malmö Breast Tomosynthesis Screening Trial reported the same density groups subjectively, using the 2003 edition of BI-RADS. In the Malmö-study and in the Oslo-trial more breast cancers were detected in DBT across all density categories [66; 96]. In the Oslo trial the results were similar also after an automated software was used to measure breast density [111]. In STORM-II, density was reported by majority score in two groups: BI-RADS 1-2 or BI-RADS 3-4 (edition 2003). Detection-rate increased for women with dense breasts after screening with DBT in STORM-II [62]. Also other studies have shown that DBT detects more cancers in dense breasts compared with DM, [111-113], but density has been measured manually in some of the studies [112; 113].

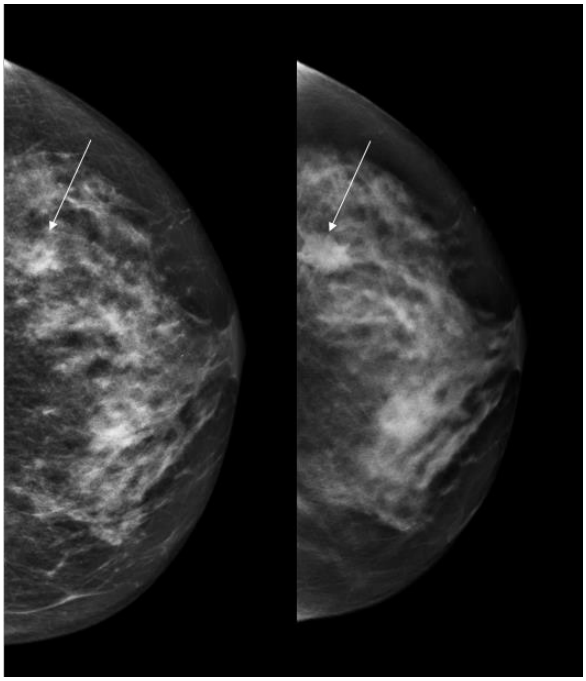

Figure 6: Synthetic 2D image (left) and 1 mm DBT-plane (right) in a craniocaudal DBT of left breast in a woman with breast density BI-RADS-c. The cancer in lateral part of the breast is difficult to identify in the SM, but more conspicuous at DBT.

Mammographic density was reported by the radiologists and not by an automated software also in the STORM-II-trial and Malmö-trial, which might be considered a limitation of the studies. Given the known substantial inter- and intra-reader variability of breast density assessment the method is not optimal, and automated methods should be preferred [107].

Notably, the effect of DBT on screening outcome of women with dense breasts should not be evaluated on detection-rate alone, but also on other performance indicators as well as long-term screening outcomes. Several studies have reported improved specificity when screening with DBT compared with DM after results have been stratified by breast density. Conant et al. found that DBT was associated with increased specificity in addition to increased cancer detection across all breast density categories in a cohort study from the US [112]. Bernardi reported improved specificity after DBT versus DM, and a larger reduction in recall rate for DBT

compared with DM in women with dense versus fatty breasts [114]. Similar results were found in the TOMMY-trial with data from UK; higher specificity for all breast densities with addition of DBT to DM [115].

To what extent the additional cancers detected after DBT is followed by a reduction in interval cancer rates in different density categories is also relevant. Importantly, an increased detection rate should be balanced against the patient burden related to additional false positive screening results and the possibility of overdiagnosis. Comparing histological grade of screen-detected breast cancers after screening with DBT versus DM is therefore considered an important early performance measure when comparing screening methods.

Before the To-Be trials started, knowledge about all aspects of performance and outcome were needed also for women with different mammographic density.

1.5 Mammographic features

The principle for breast cancer detection by mammography is the visualisation of suspicious findings consistent with pathology on the mammogram. Due to the heterogeneity of breast cancer the pathology appears with different mammographic feature; e.g. circumscribed-, indistinct-, or spiculated masses. Other breast cancers are growing like “Indian files” or in a spider web-like appearance, often difficult to recognise in a mammogram, while some are spreading in the ducts in the early in-situ stage while simultaneously calcifying (Figure 7) before they might transform into an invasive tumor. Thus, when reading a mammogram different imaging features have to be interpreted.

Several classification systems for mammographic features exist, one of the most used systems are BI-RADS.

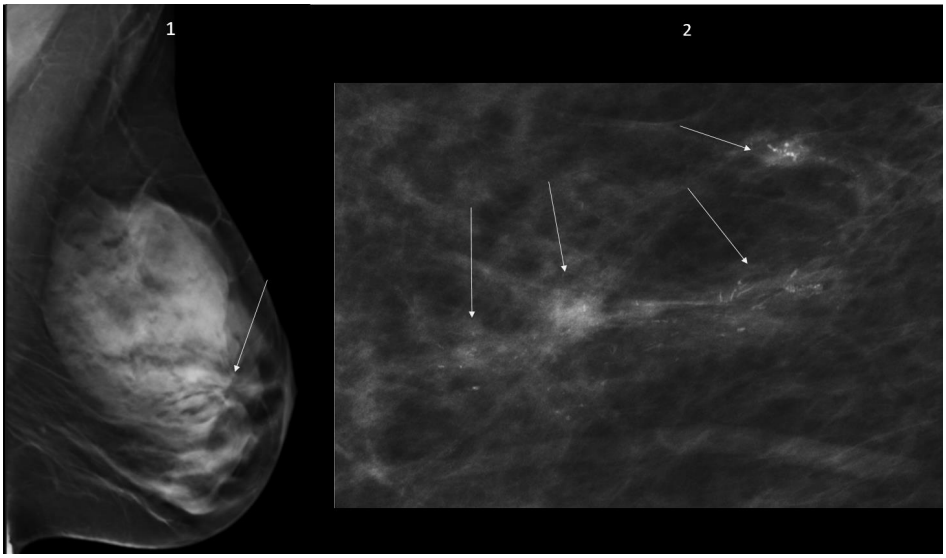

Figure 7: A DBT-plane showing an architectural distortion behind the nipple; histopathological proven invasive breast cancer (1). A DM magnification view of microcalcifications of fine linear branching type, which turned out to be ductal carcinoma in situ, van Nuys grad 3 at histology report (2).

1.5.1 Mammograpfhic features in BI-RADS

In BI-RADS [95] the mammographic features are called "*Important findings"* and are classified into following categories:

- Mass
- Architectural distortion
- Asymmetries
- Calcifications
- Associated features
- Special cases

A mass is a 3D lesion visible in two different projections and is further described according to its shape (oval, round or irregular), margins (circumscribed, obscured, microlobulated, indistinct, spiculated) and density (high, low or fat-containing). The

features can be used to estimate the probability of malignancy of the findings. For further details regarding classification the BI-RADs-atlas should be consulted.

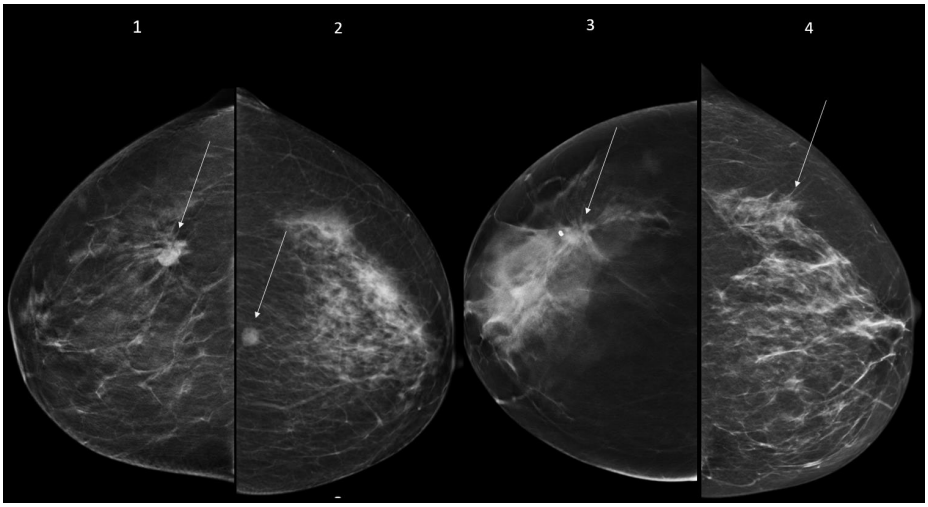

Figure 8: Mammograms from women with invasive breast cancers: A DBT-plane showing an indistinct mass (1). DM with a circumscribed mass in a fatty area of a left breast (2). A spiculated mass in a DBT plane of a right breast (3). A SM with an asymmetry the lateral part of a left breast (4).

1.5.2 Other classification systems

The classification system of Tabar et al is an alternative classification system for mammographic features [116]. He has classified mammographic features in the following categories: Stellate without calcifications, circular/oval without calcifications, powdery calcifications with or without tumor mass, crushed stone-like calcifications with or without tumor mass, casting-type calcifications with or without tumor mass and others. This means that the classification system given by Tabar focuses on the calcification feature; if calcifications is present it can be with or without a mass. This is in contrast to BI-RADS where focus is on the mass, which can be with or without calcifications.

BreastScreen Norway uses a modified version of BI-RADS for reporting mammographic features in women recalled for further assessment. In BreastScreen Norway mammographic features are classified according to the following categories: mass (circumscribed or spiculated mass), asymmetry, architectural distortion, calcifications, and associated features. In the studies included in this thesis, a classification system being more similar to BI-RADS than the current Norwegian system were used, described in detail in Chapter 3.4.

1.5.3 Associations between mammographic features and histopathology

Several studies have shown associations between mammographic features in DM and histopathological prognostic and predictive characteristics [42; 116; 117]. According to BI-RADS, a circumscribed mass is a benign finding, while microlobulated, indistinct and spiculated mass indicate susceptibility of breast cancer.

Tabar et al. has found a poor long-term prognosis in women with tumors associated with "casting-type-calcifications" (i.e. linear, or linear branching-calcifications in BI-RADS) when they were followed 24 years after invasive breast cancer diagnosis. Similar results were not found in other tumors in the same size category at mammography [116]. Other studies have confirmed spiculated masses (i.e. stellate lesions in Tabar classification system) to be associated with especially luminal A-like tumors, often low grade, with good prognosis if smaller than 15 mm [42; 44; 117]. Architectural distortion often represents low-grade cancers [118; 119].

1.5.4 Mammographic features and DBT versus DM in screening

DM has a high resolution compared with other imaging techniques in radiology and also compared with DBT and is able to present small calcifications and lesions as long as the features are not obscured by overlapping breast tissue. However, the thin planes in the DBT should visualize margins of suspicious and probably benign masses better than DM in breasts with BI-RADS category b-d. Thus, the margins of a malignant mass, which more often is spiculated, indistinct or microlobulated, should be easier to differentiate from an e.g. circumscribed benign cyst when using DBT. Several studies report architectural distortions and masses, especially spiculated

masses, to be better visualised in DBT compared with DM [66; 96; 113; 120]. Spiculated masses are the most common feature in breast cancers after screening with DM and DBT [66; 120].

Phantom studies indicate that DBT yields better performance for detecting masses than DM [37]. Presentation of mammographic features might differ between vendors and type of equipment. Using a wide angle has been shown to yield better resolution of masses compared with a narrow angle [37].

In the Oslo Tomosynthesis Screening Trial, 79% of the invasive cancers only detected by DBT, were classified as spiculated mass or architectural distortion; features that are more common in low-grade cancers compared with high grade cancers [118; 119]. The majority of the additional cancers detected by DBT had low Ki-67 in the Oslo trial. Invasive lobular carcinomas are often visible in DBT as spiculated mass or architectural distortions. The features can be difficult or impossible to detect in DM [121; 122]. Dense breasts, small tumor size, luminal-A like invasive cancer and low Ki 67 expression have been shown to be more frequent in DBT-only detected cancers [68; 123].

Masses are more conspicuous at DBT compared with DM, but there are some concerns regarding calcifications in DBT [37; 124]. Even though calcifications are associated with DCIS, the results from the studies of Tabar et al. and others, indicate that calcifications sometimes can be crucial to perceive and recall. When comparing the thin planes in a DBT versus DM, similar results have been reported regarding evaluation of calcifications [37]. However, in SM calcifications present differently from that in DM; larger calcifications are highlighted in SM and therefore can give the impression of equal or better visualization compared with DM [37; 124]. The concern related to calcifications and DBT is first and foremost present for subtle calcifications, where a study has reported reduced detection with SM versus DM [37]. Thus, in a minority of cases, calcifications have been classified differently in DBT compared with DM, which might have a clinical relevance, especially if casting

or fine linear branching calcifications are missed at DBT, but possible to percept at DM [116; 125; 126].

Further, it has not been studied in enough detail if the possible differences between vendors have any clinical relevance when it comes to presentation of calcifications in DBT including SM [36; 37].

Thus, before start-up of the To-Be trials, there were substantial knowledge gaps related to use of DBT in a screening setting. Limited knowledge existed on whether the mammographic features were similar when using different screening techniques and systems, or whether tumors detected after screening with DBT and DM had similar histopathologic characteristics. It was questionable if the radiologists were able to perceive all the features and interpret all the information in a DBT-uptake, and if the quality of synthetic 2D images and thin planes was sufficient for the perception and characterisation of tiny calcifications and other lesions in different systems. To explore these issues, the different features visible at mammography should be defined and the results compared between DBT and DM. Further, studies performed with different DBT-systems should be compared. In 2015, high quality RCTs were needed to answer some of the questions.

Search for studies relevant for the thesis was performed until May 2022.

2. AIMS OF THE THESIS

The thesis was based on results from To-Be1, a randomized controlled trial (RCT) aimed to prove superiority effect of DBT versus DM in an everyday screening setting in Bergen [127].

The aims of the thesis was to investigate

a) time spent on screen reading and consensus, mean glandular dose, number of cases discussed at consensus and recall due to mammographic findings for DBT versus DM (Study 1).
b) recall, false positive screening results, screen-detected cancers, positive predictive value of recall and biopsy, and histopathological tumor characteristics for DBT versus DM stratified by mammographic density measured by an automated software (Study 2).
c) distribution of mammographic features in women recalled after screening with DBT versus DM and to explore associations between mammographic features and final outcome of the screening examination (Study 3).

3. MATERIALS AND METHODS

3.1 Study sample and data collection

All studies included in this thesis comprised data from the To-Be1 trial, an RCT comparing DBT versus DM in Bergen, as a part of BreastScreen Norway, during the period from 2016 to 2017 (NCT02835625). The trial was approved by the Regional Committee for Medical and Health Research Ethics in the South East of Norway (2015/424) [127] and headed by the Cancer Registry of Norway. The To-Be1 trial was supported by the Research Council of Norway (project number 247941/H10). We received pseudonymized data from the Cancer Registry of Norway. When To-Be1 ended, To-Be2 started. After replacing the GE Seno Claire equipment with Pristina, all women attending the screening unit in Bergen were offered participation in To-Be2 where they were screened with DBT. Data and results from To-Be2 are not a part of the studies included in this thesis.

Study population

A total of 44,266 women aged 50-69 were invited to screening in Bergen during the study-period, whereas 32,976 (74.5%) attended. The women were informed about the trial in the invitation letter and asked to arrive at the screening unit in time to get more information. All attending women were invited to participate in the trial after oral and written information, except those with implants (due to concerns about radiation-dose). Written informed consent was obtained before randomization to two-view DBT (including SM), or two-view DM. In total 29,453 women (89.3% of the attending women) accepted the invitation to participate in To-Be1 during the study period, two years. A total of 354 women in the DBT-arm and 350 in the DM-arm were excluded in the analyses because of previous breast cancer, breast symptoms or metastases from melanoma, leaving 14,380 women screened with DBT and 14,369 screened with DM as the study population (Figure 9).

At the screening unit, three radiographers participated during screening examinations; one interviewed the women, and two performed the acquisition of the mammograms.

Equipment

Equipment from GE (Seno Claire 3D Breast Tomosynthesis TM) was used. During the acquisition of a DBT the x-ray-tube moved 25 degrees. Nine low-dose images were reconstructed into 1 mm planes + 10 mm slabs + a SM. An automated software (VolparaDensity, version 1.5.1) was integrated in the picture archiving and communication system. All examinations were analysed by the automated software Volpara, which uses information from each pixel and information from the DICOM-header and thereby reports data of breast volume, fibroglandular volume, compressed breast thickness, volumetric breast density (%) and VDG 1-4, as well as BMI. For BMI, data was most frequently extracted from a questionnaire used in BreastScreen Norway from 2006-2016, however, breast volume was used as a proxy for BMI in women without available weight and height data [128].

Hanging protocol

The hanging protocol, i.e. how new and prior mammograms were presented for the radiologists, presented two-views for DM and two-views of SM and planes for DBT and included priors if present. Up to 5 prior DM examinations were available at the workstation, but all priors were used only for challenging cases and were not part of the hanging protocol. The hanging protocol included two sets of prior DM screening examinations, commonly representing examinations 2 and 4 years before the actual examination. The workstation allowed presentation of the previous and older priors, but not only the 4-years-old images as would have been preferred. The radiologists however, as a main rule, skipped comparing the 2-years-old images, and compared the DM or SM and planes with the 4-years-old images if an interesting feature appeared. The DBT-slabs were used only in special cases.

Screen reading

DBT and DM examinations were read at IDI 5 MP monitors by eight breast radiologists with different screening experience (0-19 years). The radiologists had no DBT screening experience, other than an 8-week long pilot period with 300 DBT examinations that was performed before start of the trial. Furthermore, the radiologists participated in a training session before reading DBT in To-Be1, and

DBT had been used for assessment after screening and in symptomatically referred patients in approximately one year.

Each radiologist was supposed to read an equal number of DBT and DM according to the established roster. However, the plan was not strictly followed, because of varying work speed of the readers and the fact that the trial was run in a busy daily practice, a pragmatic trial.

In the trial, mammograms were read using independent double reading with consensus, the standard procedure used in BreastScreen Norway, including a 5-point interpretation scale. A score of 1 indicated no recall was necessary. Score 2 indicated a probably benign finding, score 3 indicated intermediate suspicion of malignancy, score 4 indicated a probably malignant finding and score 5 indicated a malignant finding at mammography. All mammograms with a score of 2 or higher from at least one radiologist were discussed at a consensus meeting, where at least two radiologists participated. If at least one radiologist had given a score 3 or higher, the woman were to be recalled for further assessment, while examinations with a score of 2 was discussed at consensus whether to be recalled or not.

Due to restrictions described in the RCT-protocol [127], no feedback on recall or detection rate was given to the radiologists during the trial, except if quality parameters was outside usual accepted levels in BreastScreen Norway. This happened once, after 18 months when the recall rate was below 2%.

The screening history was defined as prevalently- or subsequently screened, defined as attending screening for the first or subsequent time in BreastScreen Norway. Consensus and recall were defined as number of women discussed in consensus/recalled because of a mammographic finding. A false positive screening result was defined as recall due to a mammographic finding with no cancer diagnosed at assessment. Positive predictive value (PPV) of recall and biopsies was defined as number of women with breast cancer among those recalled and biopsied, respectively. Breast cancer was defined as invasive cancer or DCIS (except in the presentation of subgroups which only affects invasive breast cancers).

Information from histology reports about histopathological characteristics of invasive cancers were given by tumor diameter, histologic grade (1-3), lymph node involvement (+/-) and subtypes classified into five subgroups based on immunohistochemistry after the consensus in St.Gallen 2013 [41]. Low Ki67-level was defined as Ki-67 level <30%, high level as Ki-67 ≥30%). Information on DCIS (diameter and grade according to van Nuys classification system) was collected from histology reports.

Automated software (Volpara) used information from raw image data and data extracted from the DICOM-header to calculate measures of mean glandular dose (MGD) of each exposure. The sum of the radiation doses for both views and breasts divided by two was calculated by the software and presented as average MGD per examination.

At consensus the radiologists classified the mammographic features according to a modified BI-RADs scale. The features included circumscribed mass, obscured mass, indistinct mass, spiculated mass, architectural distortion, asymmetry, calcifications, associated features and mass with calcifications. All features were defined according to BI-RADS, except for masses including calcifications. Microlobulated mass was recorded as indistinct mass.

Assessment

Further assessment was performed by the same eight radiologists that performed the screen reading. Additional imaging and ultrasound with clinical examination was standard procedure for assessment. Based on the findings, the radiologist decided whether a needle biopsy and/or further imaging as MRI or contrast enhanced mammography were needed.

Biopsies were performed under ultrasound or stereotactic guidance. DBT-guided biopsies were not available. All information regarding the screening examinations were reported electronically by the radiologists to the Cancer Registry of Norway, as a part of the standard procedures in the screening program.

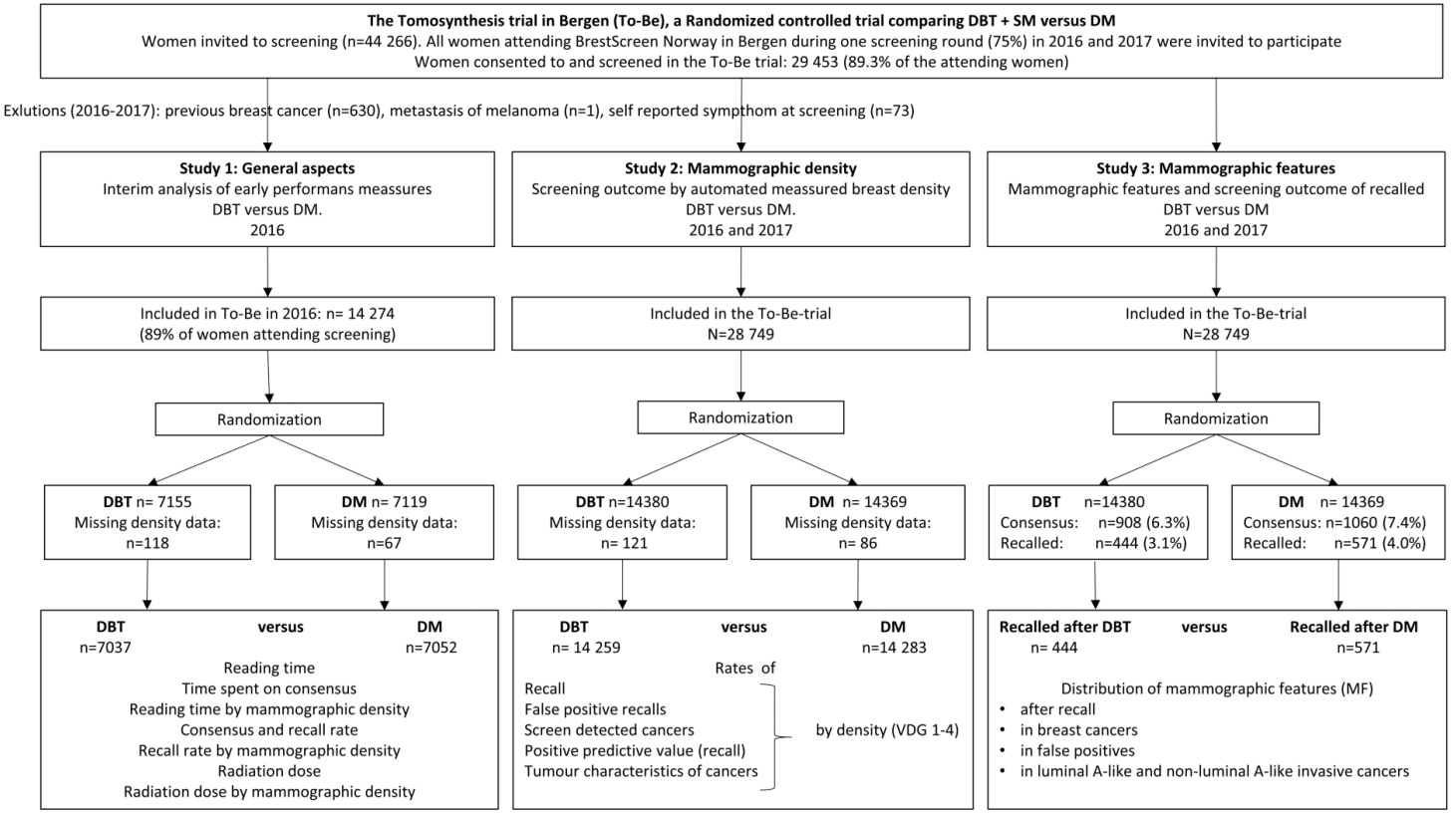

Figure 9: Profile of the Tomosynthesis trial in Bergen (2016-2017), and of studies included in the thesis.

3.2 Study 1 (Interim analysis after the first year)

Study 1 included data from 7155 women screened with DBT and 7119 women screened with DM in 2016 (Figure 9). Because of missing data on density, 185 participants were excluded. The study includes interim analysis after the first year of the To-Be1 trial. The paper described the RCT and presented time spent on initial screen reading and on consensus. MGD was described and compared by breast density (VDG1-4) for DBT and DM. Further, recall and consensus were compared for DBT and DM. The age of participants were presented as mean, median and as distribution of age groups in DBT and DM.

We described the mean and median values for time spent on the acquisition process (min:sec) – from the time the woman entered the screening room until she left. Time was measured using a stopwatch for 438 randomly selected women randomized to have DBT and for 534 women screened with DM in March 2017. Mean end median time spent on screen reading and consensus were measured electronically for each participating woman from the moment when the radiologist entered the case (and the mammogram) in the screening database until the result was reported and the radiologist switched to the next woman. Cases were the reading time exceeded 10 minutes on initial screen reading, and 15 minutes on consensus were excluded, because we assumed the radiologists had been interrupted. Results regarding time were stratified by screening history, time since start of trial, and mammographic density (VDG 1-4).

Number of examinations discussed at consensus and recalled were compared for DBT versus DM and stratified by screening history, by time since start of trial and by breast density. Risk ratio for consensus and recall adjusted for mammographic density were calculated for DBT using DM as the reference.

In supplementary material, screening-experience of participating radiologists was presented (month and year when starting screen-reading in BreastScreen Norway and DBT in To-Be1). Number of DM screens for each radiologist before start-up of To-Be1, number of screen-reads (DBT and DM) in To-Be1, rate discussed in consensus

and rate recalled after DBT and DM were reported. Further, cumulative number of DBT and DM screen readings and subsequent rates of consensus and recall were presented.

Descriptive results were presented for DBT and DM separately. We used STATA software (version 15; Texas, USA) for all statistical analyses and tested difference across categories for statistical significance using two sample t-tests, Chi-square tests and ANOVA and tests of proportions (z test). A negative binomial regression model was used to test trends in consensus and recall rates according to reading volume and to estimate risk ratio (RR) and 95% CI of consensus and recall rate for DBT using DM as a reference. Both crude RRs and RRs adjusted for mammographic density were calculated. Covariates in the adjusted models included breast density and an interaction effect between the screening technique and density. A p-value of less than 0.05 was considered statistically significant.

3.3 Study 2 (Mammographic density)

In study 2, data from the entire study period (2016-2017) were used (Figure 9). In total 14 380 women were screened with DBT and 14 369 women screened with DM. We excluded 121 DBT-screened and 86 DM-screened women because of no information of density. We compared recalls, false positive screening results, biopsy, cancer detection, positive predictive value of recalls and biopsies, and histopathologic tumor characteristics stratified by mammographic density, among women screened with DBT versus DM.

Baseline characteristics of the women were presented with median and interquartile range (unless otherwise indicated) for women screened with DBT and DM and included: Age, screening history (n, %), BMI, breast volume, fibroglandular volume, compressed breast thickness, volumetric breast density. Further, the distribution of breast density was presented as Volpara Density Grade (VDG) 1-4 (n, %) and by volumetric breast density quintiles (number and 95% CI) for women screened with DBT and DM separately [105]. In addition to exclusions due to prior breast cancer,

symptoms or metastases (354 for DBT and 350 for DM), 121 women were excluded in the DBT arm and 86 in the DM arm because of missing density values.

Recall, false positive, biopsy, screen-detected cancers, and PPVs were presented by VDG 1-4 for DBT and DM including 95% CIs around proportions and by Z tests. We used Chi-square to test differences in categorical distributions.

Tumor diameter (mean), distribution of histologic grade (%), positive/negative lymph node status (%) and immunohistochemically subtypes (%) for invasive breast cancers were presented for DBT and DM separately. Histopathologic tumor characteristics were presented as percentages of women with invasive breast cancer.

Log-binominal regression models were used to analyse relative risk (RR) of recall, false positive, and screen-detected cancers for DBT and DM by VDG. We adjusted for age groups, screening history, and by breast volume (continuous). We modelled the interaction between VDG and the screening technique using DM and VDG1 as a reference. A p-value lower than 0.05 was defined as statistically significant. We used STATA software 16 and R software (version 3.6.1; Vienna, Austria).

3.4 Study 3 (Mammographic features)

In Study 3, we used data from women recalled due to a mammographic finding after screening in the To-Be1 trial (2016-2017); 444 screening examinations with DBT and 571 with DM [127]. We compared the distribution of mammographic features in recalled women after DBT versus DM. Furthermore, we explored the distribution of recalled women with breast cancer and women with false positive screening results, for DBT versus DM, and lastly we compared the distribution of mammographic features in subgroups of invasive cancers by screening technique.

Pseudonymized data from the Cancer Registry of Norway included information about the assessment, mammographic features and histopathological findings. A modified classification system similar to the BI-RADS 5th edition was used (Table 7). Subgroups of invasive cancers were presented as either Luminal A-like or non-luminal A-like (Luminal B HER2-, Luminal B HER2+, HER2+, and Triple negative).

Table 7: Overview over mammographic features in the To-Be1 trial and in BI-RADS.

Term in our study	**Term in BI-RADS**	**Comment**
Circumscribed mass	Circumscribed mass	
Obscured mass	Obscured mass	
Indistinct mass	Indistinct mass	
Indistinct mass	Microlobulated mass	
Spiculated mass	Spiculated mass	
Architectural distortion	Architectural distortion	
Asymmetries	Asymmetries	
Calcifications	Calcifications	
Associated features	Associated features	
(Not used)	Special cases	In BI-RADS: intramammary lymph node or a wart on the skin are examples. In To-Be1: one of the other categories were used (often circumscribed mass).
Mass with calcification	(Not used)	In BI-RADS: classified as one of the masses, depending of margin

We used descriptive statistics to present mean age with standard deviation (SD), while screening history, BI-RADS density, assessments used at recall (ultrasound or ultrasound as well as other imaging), biopsies performed, and mammographic features were presented as numbers and percentages with 95% CIs among the recalled women for DBT and DM separately. Numbers and percentages with 95% CI were also used when describing mammographic features in recalled women with malignant or benign outcome (defined as benign biopsy or biopsy not performed after

recall), separately for DBT and DM, with different denominators: a) by recalled women with malignant and benign outcome for DBT and DM, b) by each feature for DBT and DM, and c) by number of screened women for DBT and DM. Further, the distribution of mammographic features for luminal A-like and non-luminal A-like cancers were presented by subgroups and by each mammographic feature for DBT and DM separately. We tested for differences using 95% CIs. STATA software (version 15; Texas, USA) was used for all data analyses.

4. RESULTS

4.1 Study 1

Age, screening history and mammographic density did not differ between women screened with DBT versus DM after the first year of the To-Be1 trial. The radiologists spent in average 1:11 min:sec on screen-reading and 3:12 min:sec on consensus for DBT and 0:41 min:sec at screen-reading and 2:08 min:sec at consensus for DM, ($p<0.01$ for both) (Table 8).

Table 8: Mean time (min:sec) spent on screen reading per radiologist and consensus for DBT screened women in the To-Be1 trial (2016-2017)

	Reading time DBT (min:sec)			Time spent at consensus DBT (min:sec)		
	N	Mean	SD	N	Mean	SD
2016	7155	1:11	0:42	460	3:12	1:57
2017	7579	1:00	0:35	510	2:30	1:30

Additional analyses showed that time spent on initial interpretation and on consensus decreased during the second year (2017) of To-Be1 for DBT (Table 8). For DBT, reading time was 1:10 min:sec for prevalently screened and 1:11 min:sec for subsequently screened ($p<0.01$). For DM the reading time was 0:33 min:sec for prevalently screened, and 0:43 min:sec for subsequently screened, $p<0.01$. Fewer cases were discussed at consensus and recalled after DBT versus DM (Table 9).

Table 9: Percentages (%) of screening examinations discussed at consensus and recalled for DBT versus DM in the To-Be1 trial in Bergen, 2016.

	DBT (n=7037)	DM (n = 7052)	p *
	% (95% CI)	% (95% CI)	
Discussed at consensus	6.4% (5.8–7.0)	7.4% (6.8–8.0)	0.03
Recalled	3.0% (2.6–3.4)	3.6% (3.2–4.0)	0.03

*ANOVA

No statistically significant difference in MGD per examination was found for DBT (2.96 mGy) and DM (2.95 mGy) ($p=0.433$).

4.2 Study 2

For women with VDG 1, 2.1% were recalled after screening with DBT and 3.3% after screening with DM (p=0.001). For women with VDG 2, recall rate was 3.2% for DBT versus 4.3% for DM (p=0.002). There was no difference in recall rate between DBT and DM for those with VDG 3 and VDG 4.

The rate of false positive screening result in the VDG 1 group was 1.6% for DBT compared with 2.8% for DM (p<0.001). The result was 2.4% of women in VDG 2 after screening with DBT versus 3.6% after DM (p<0.001). No statistically differences were observed for women in VDG 3 and 4.

We found no statistical difference between rates of screen-detected cancers stratified by density-groups when comparing DBT with DM.

DBT resulted in a higher PPV for screen-detected cancer in VDG 2 compared with women with VDG 2 in DM; 24.0% for DBT versus 14.6% for DM (p=0.01).

The adjusted relative risk of recall increased by density categories for DBT, but not for DM (Table 11). There was an increased adjusted relative risk of false positive screening result in women with VDG 2 and 3 for DBT, but no difference was observed for DM. Further, the adjusted relative risk of screen-detected cancer increased for higher density grades for DBT, but no such difference was observed for DM (Table 10).

Table 10: Crude and Adjusted Relative Risk (RR) with 95% CI of recall, false positive screening result and screen-detected cancers for 14 259 women screened with DBT and 14 283 women DM, by mammographic density, Volumetric Density Grade (VDG) 1–4.

	RR of recall for DBT				**RR of recall for DM**			
VDG	Crude	95% CI	Adjusted	95% CI	Crude	95% CI	Adjusted	95% CI
1	1.0	…	1.0	…	1.0	…	1.0	…
2	1.6	1.2 - 2.0	1.8	1.4 - 2.4	1.3	1.0 - 1.6	1.3	1.0 - 1.6
3	2.0	1.5 - 2.6	2.4	1.7 - 3.3	1.2	1.0 - 1.6	1.1	0.8 - 1.5
4	1.5	1.0 - 2.3	1.8	1.1 - 2.9	1.2	0.9 - 1.7	1.1	0.7 - 1.6
	RR of false positive screening result for DBT				**RR of false positive screening result for DM**			
VDG	Crude	95% CI	Adjusted	95% CI	Crude	95% CI	Adjusted	95% CI
1	1.0	…	1.0	…	1.0	…	1.0	…
2	1.5	1.1 - 2.0	1.7	1.2 - 2.3	1.3	1.0 - 1.6	1.2	0.9 - 1.6
3	2.1	1.5 - 2.9	2.3	1.6 - 3.2	1.2	0.9 - 1.5	1.0	0.7 - 1.4
4	1.6	1.0 - 2.5	1.6	0.9 - 2.7	1.2	0.8 - 1.8	1.0	0.6 - 1.5
	RR of screen-detected cancer for DBT				**RR of screen-detected cancer for DM**			
VDG	Crude	95% CI	Adjusted	95% CI	Crude	95% CI	Adjusted	95% CI
1	1.0	…	1.0	…	1.0	…	1.0	…
2	1.7	1.0 - 2.9	2.3	1.3 - 4.2	1.3	0.7 - 2.4	1.7	0.9 - 3.1
3	1.6	0.9 - 3.0	2.8	1.3 - 5.7	1.5	0.8 - 2.8	2.1	1.0 - 4.6
4	1.4	0.5 - 3.4	2.8	1.0 - 8.0	1.3	0.5 - 3.2	2.2	0.8 - 6.2

4.3 Study 3

Asymmetry was the most common mammographic feature among recalled women screened with DBT as well as with DM (Figure 10). We found 108 women recalled because of asymmetry after DBT, and 222 after DM (p<0.001).

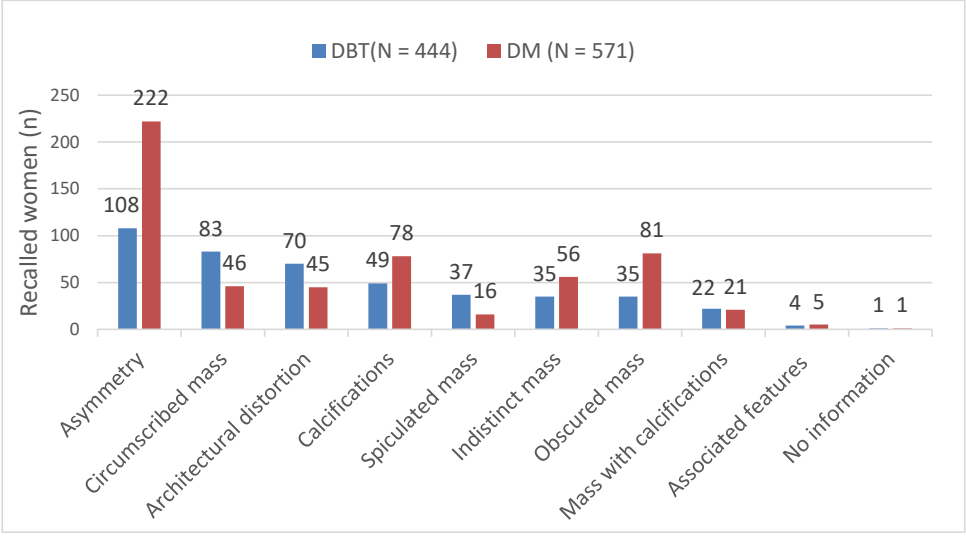

Figure 10: Distribution (n) of mammographic features of recalled women after screening with DBT versus DM in the To-Be1 trial in Bergen, 2016-2017.

Spiculated mass was the most frequent feature among breast cancer detected by DBT, while calcification was the dominant feature for DM (Figure 11). Indistinct mass was the second most frequent mammographic feature among breast cancers both for DBT and DM. The proportion of luminal A-like cancers did not differ statistically between DBT (58.7% (44/75, 95% CI 46.7-69.9)) and DM (61.4% (43/70, 95% CI 49.0-72.8)). Among non-luminal A-like invasive breast cancers, spiculated mass was the dominant feature after screening with DBT and DM.

In the features asymmetry, indistinct mass and obscured mass the number of women with a false positive screening result was 101, 19 and 34 after DBT versus 210, 40 and 78 in DM, p <0.01 for all (Figure 12).

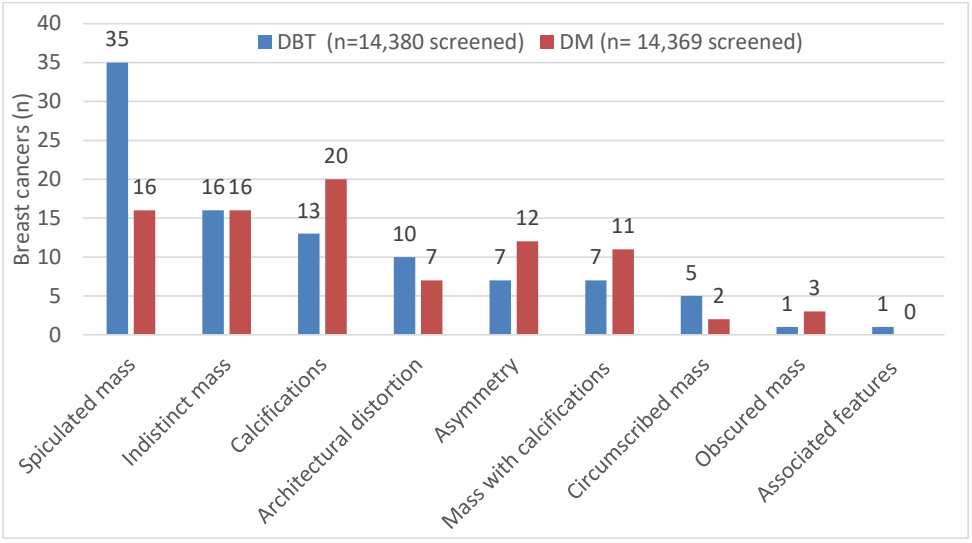

Figure 11: Distribution (n) of mammographic features of breast cancers (invasive or ductal carcinoma in situ) after screening with DBT versus DM in the To-Be1 trial, 2016-2017.

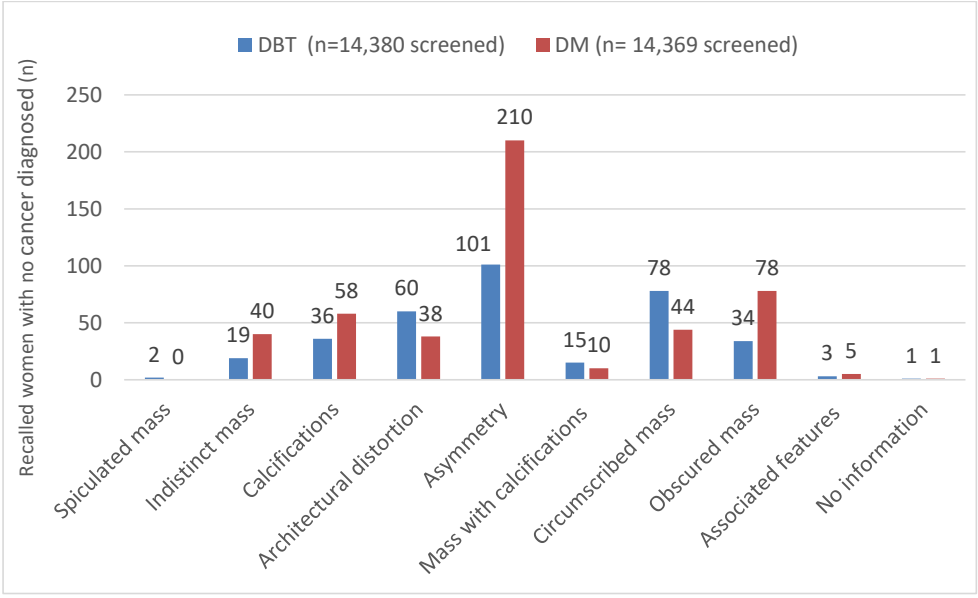

Figure 12: Distribution (n) of mammographic features of women recalled because of mammographic finding but not diagnosed with breast cancer after assessment after screening with DBT versus DM in the To-Be1 trial, 2016-2017.

5. DISCUSSION

5.1 Main findings

Interim analyses after one year of the ToBe1 trial showed longer examination- and reading-time and less recalls for women screened with DBT versus DM. Radiation dose did not differ statistically between the two techniques.

Lower rates of recall and false positive screening results was found among women with low mammographic density and screened with DBT versus DM. Adjusted relative risk of recall and screen-detected cancer increased by mammographic density for women screened with DBT but not for women screened with DM.

The distribution of mammographic features differed statistically between women recalled after DBT versus DM. More women were recalled and diagnosed with breast cancer because of spiculated mass after DBT versus DM. Asymmetry, indistinct and obscured mass was less frequent in women with a false positive screening result after screening with DBT versus DM.

5.1.1 Time spent on screen reading and consensus

Radiologists spent on average 30 seconds more reading DBT compared with DM. The additional time spent on screen reading is in line with other prospective European screening trials [63; 64; 66; 76]. Time spent on initial screen-reading is affected by number of views (one or two-view mammography), procedures related to the hanging protocol, the individual radiologists, the reporting systems and reading infrastructure and computer platforms.

A total of 30 seconds longer reading time for each woman may seem insignificant when considering the feasibility of using DBT in screening. If a radiologist reads 120 mammograms a day, this implies an extra hour reading time each day, or five hours a week. Notably, some radiologists seem to be fast readers and other slow readers [76], thus, the extra time needed for DBT in screening is expected to vary between radiologists.

Reading time for DBT decreased during our trial from 1:11 min:sec (SD 0:42) in 2016 to 1:00 min:sec (SD 0:35) during the second year of To-Be1 (2017). Unpublished quality assurance results from To-Be2 showed that time spent on initial screen reading for DBT further decreased to 0:53 min:sec (SD 0:36) in 2018 and 0:48 min:sec (SD 0:34) in 2019. Time spent on consensus also decreased during the To-Be trials, from 3:12 min:sec in 2016 (SD 1:57) to 2:26 in 2019 (SD 1:18). This indicates that time needed might decrease with increased experience, as has been reported also in another study [76].

In To-Be1 we measured time spent on screen-reading for each DBT and each DM examination. Whether radiologists needed to take a break more often when reading DBT compared with DM was not registered.

For screen reading, outliers above 10 minutes were excluded, whether this was the better choice for exclusions is not obvious. Also when reading times exceed 5 minutes, they are likely to be caused by interruptions, and thus not reflect true reading times.

The over-all time spent on screen-reading is affected by IT-issues; two-view DBT-images are significantly larger compared with DM (in our study, DBT most often were 500-3000 MB compared with 60-80 MB for DM). Thus, it is important to optimize both storage, eventual prefetching of images, power and speed of IT systems before introducing DBT in screening.

The hanging protocol used in To-Be1 might seem comprehensive. However, when reading DBT the radiologists had to pay attention to the DBT-screen, and only if something appeared as suspicious, priors were of interest and consulted. The marginal increase of only 1 second in mean and median reading time in subsequently screened women in DBT compared with prevalently screened, supports this assumption, and might explain the rather short / similar reading time in our study compared with other studies [63; 64; 66; 76], in spite of several priors mammograms available in To-Be1.

If artificial intelligence (AI)-guided diagnostic software are introduced in screening, reading time will probably be affected. It is, however, difficult to know how AI will impact reading time. If these tools are used as a support to the radiologist, some extra seconds may be needed. However, if more breast cancers are detected with subsequent fewer interval cancers, this extra time might be justifiable. If AI replaces one or both of the radiologists in a substantial part of screen-readings, reading volume might be substantially reduced, which is especially important if screening is performed with DBT.

5.1.2 Radiation dose

Radiation of the breast is associated with increased risk of cancer [129]. Even though mammography is known to have low doses, it is important to keep doses as low as possible, particularly in population-based screening [80]. Thus, our finding of no statistical difference in mean glandular dose (MGD) among women screened with DBT versus DM is positive. The results stratified by breast density are even more uplifting; 95% CI for MGD were overlapping in all density categories in DBT versus DM. These findings are different from that of Gennaro et al reporting a significant increase (mean: 38% increase) of automatically (Volpara v.1.5.2.0) measured MGD for DBT-screened using mammography equipment from a different vendor [79]. The same study found that the incremental DBT-dose increased in women by increasing compressed breast thickness. Breast thickness is inversely correlated to breast density [130]. Gennaro et al also found that MGD was influenced by breast density when using their equipment; dose decreased with decreasing density for DBT and DM. In the equipment used in our study, doses were set to be equal in DBT and DM, further, we found no difference in doses between different breast densities. If a higher dose is needed in DBT compared with DM to detect pathology in dense and in thick breasts, it is possible that the dose utilized in the present trial was too low for optimal detection of lesions in dense breasts. However, our study is not designed to answer this.

Other studies have machine-reported doses, where glandular fraction of the breasts are often assumed to be 50% [64; 66; 79]. Based on this assumption the vendors

report dose data stored in the DICOM header. These dose-data were different from those derived from the software for automated density assessment as used in our study, where patient-specific dose was used. In the Oslo-trial, using the same equipment as Gennaro et al, a 23% increase in radiation dose was found (machine-reported) for DBT versus DM [66], while the Malmö trial reported a MGD of 1.6 mGy for DBT and 1.2 mGy for DM, a 33% increase in radiation dose for DBT per view [64]. In the Malmö trial, one view DBT was used. The reason for different results in Gennaro and the Oslo-trial could be due to the different systems for dose-reporting, but also the differences in compressed breast thickness in the two studies.

5.1.3 Recall

Fewer women were recalled after screening with DBT compared with DM after one year, and during the whole study period. These findings contrast results from a meta-analysis reporting an average of 0.5% increase in recall rate after DBT in prospective paired trials from Europe [54], where women had both DM and DBT and recall- and detection rates were compared at an individual level. Several of the included studies used sequential reading and performed an "either positive" recall rule (participants were recalled if positive result after DM or DBT). Other paired studies used arbitration after "either positive" screen reading, in which comparison of rates are hampered. Other European studies however, found similar or lower recall rate after DBT versus DM [113; 131]. Recall rate in the US is known to be 3-5 times higher compared with Europe and a decrease in recall rates for DBT versus DM is thus expected [54]. Thus, the discrepant findings in different studies and from different parts of the world might be due to methodological issues and this variation seems reasonable when the original level of recall is taken into consideration [54].

The lower recall rate among DBT-screened in To-Be1 trial (3.1%) compared with the Oslo (3.7%), Malmö (3.8%), STORM-I (4.3%) and STORM-II (4.8%)-trials [62-65] might also be explained by the limited experience in screen reading DBT among the radiologists in Bergen. One of the first European prospective studies also had slightly lower recall rate after DBT [63]. When unexperienced radiologists started screen-reading with DBT they might assume that breast cancers were visualized very

obviously, if present. This assumption might have resulted in missed cases, later detected as interval or screen-detected cancer in the next round. Retrospective reviews are needed to explore this.

If a suspicious finding was detected in To-Be1, often several priors were consulted. The radiologists assumed that if the finding also was present in one or several priors, the probability of an aggressive malignant tumor substantially decreased. Thus, the quite extensive number of available priors in To-Be1 might not be disadvantageous if primarily aiming for detecting high grade aggressive breast cancers and reduce interval cancer rate. However, if aiming to detect as many breast cancers as possible, it might also be considered a disadvantage; a slow growing breast cancer could have been present also in priors and could have been dismissed in screen reading or consensus when priors were consulted.

The fact that no results from the trial were communicated from the principal investigator to the participating radiologists during the study might have affected recall rate. If radiologists had been informed of the lower recall- and the non-superior detection-rate for DBT compared with DM earlier, this could have increased the recall rate in the DBT-arm. However, expected and acceptable ranges of values for reporting deviations to the radiologists were set before start-up of the study. A recall rate below the lower limit was reached in the last part of the study, after informing radiologists, the recall and cancer detection rate increased [127].

The radiologists were informed about detailed results of the study after it ended. The same radiologists continued reading DBT in To-Be2, where all women were screened with DBT. The trial is described elsewhere [132]. In To-Be2, the equipment Seno Claire was replaced by Pristina, because of ergonomic arguments claimed by the radiographers. With the same radiologists, recall rate increased from 3.1/1000 in To-Be1 to 3.9/1000 in To-Be2 for DBT-screened, while detection rate increased from 6.6/1000 in To-Be1 to 8.6/1000 screened in To-Be2. For women screened with DM in To-Be1, detection rate increased from 6.0/1000 in To-Be1, to 9.1/1000 in To-Be2 [132]. There might be several reasons for the higher recall rate in To-Be2 and other

DBT screening trials versus To-Be1; in To-Be2, the radiologists had still no feedback of the recall rate, however they were informed about the recall and cancer detection rate in To-Be1. This information clarified for participating radiologists that DBT does not always depict cancers very clearly, and further; suspicious findings at DBT which had not changed compared with prior DM images, could still be a slow growing cancer. Further, the equipment had been changed.

If all suspicious findings in DBT had been recalled in To-Be1, even if no change when compared with priors, the detection rate would probably have been higher, however there is also a risk of increasing level of overdiagnosis with such a strategy. Thus, the quite extensive number of available priors in To-Be1 might partly explain the lack of additional tumors with favourable histopathological characteristics compared with other DBT screening trials [127].

The risk ratio of interval breast cancer after To-Be1 has been reported to be 0.69 for DBT versus DM screened, and risk ratio for subsequent round screen-detected breast cancer 0.89 for DBT using DM as a reference in To-Be2 [132]. Although interval cancers among DBT screened were only 2/3 of DM screened in the To-Be1 trial, there were no statistical difference. Still, the editorial in Radiology stated the results as quite encouraging, and concluded that the results provided a further argument in favour of DBT as a screening tool [132; 133]. In To-Be1, the radiologists used prior mammograms in the interpretation and it was mainly women with a new suspicious finding that were recalled. More cancers were detected in To-Be2, suspicious findings were recalled even if unchanged compared with priors. However, the unanswered question is to what extent also level of overdiagnosis has increased in To-Be2 compared with To-Be1. How priors are to be used in the future to balance in the most optimal way benefits and harms of screening, is still not studied in enough detail. The results in the To-Be trials might be used as an argument for further studies, with larger sample sizes, to obtain significant results regarding the superiority of DBT before DM in screening. Several large scale DBT studies are ongoing, for instance in the United Kingdom, in Germany and in the US [134-136].

Before To-Be1 stated, the radiologists had no experience in screen reading with DBT, but the method had been available at the breast centre for a year for problem solving cases. Not all the radiologists participated in the pilot of eight weeks before the start-up of To-Be1, but all had a training session. This training period using the new method could in retrospect have been longer and should probably have been extended to at least 6 months to prepare the radiologists better. The radiologists may have been early in their learning-curve for DBT during To-Be1.

Several studies have shown an association between recall rate and detection rate [137; 138]. In To-Be1 the radiation dose, recall- and detection rates were continuously followed and presented for the steering group of the trial to ensure the trial was conducted safely and that results were within limits. In retrospect it might be interesting to consider whether continuous information also to participating radiologists about recall rate and cancer detection could have led to an increase in recall and cancer detection in To-Be1. However, the risk of bias would increase and the principle of an RCT violated [56].

The study was performed in an every-day setting, with a high work-load on participating employees at the breast centre. The number of radiologists was small in periods of the trial, which might have affected the recall and cancer detection rate. Alternative explanations of fewer breast cancers in To-Be1 compared with To-Be2 and other prospective European DBT-trials could be differences in equipment.

However, the optimal recall rate after screening with DBT as well as DM is difficult to define. It is important to bear in mind that the aim of breast screening is not to detect the highest level possible of breast cancers, but to decrease breast cancer mortality.

5.1.4 Mammographic density

We observed an increase in adjusted relative risk of screen-detected cancer in mammographic dense versus fatty breasts after screening with DBT in our study, using an automated software for density measurement. These results are in line with findings in other studies [107; 111]. We found a lower recall rate and false positive

rate for DBT versus DM in women with lower breast density (VDG 1-2), still, we observed an equal rate of screen-detected cancers for women with VDG 1, and a non-significant difference for VDG 2 (with 9 more breast cancers detected, of whom 5 invasive, in VDG 2 after DBT compared with DM). This indicates DBT to be superior to DM in VDG 1-2 in our study. Further, the majority of women (69%) had lower breast density (VDG 1-2).

The rate of screen-detected cancer did not differ for DBT versus DM among women with density categories VDG 3-4. Given the known increased risk of screen-detected as well as interval breast cancer in women with dense breasts, it is possible that supplemental screening will be recommended for women with the densest breasts in the future [139]. Neither the To-Be1-trial nor other studies have found the increase in detection rate for women with dense breast as high as expected after DBT.

The fact that neither DM nor DBT are found to have as high sensitivity in dense as in fatty breasts, is worth noting when discussing whether to introduce screening in also older- and younger age groups than 50-69 years [49; 140]. Perhaps there is a need for personalized supplementary screening in a higher proportion of women in the youngest age groups if screened, since mammographic density are known to be higher. The method could be MRI or other methods (i.e. ultrasound ot contrast enhanced mammography) [139].

5.1.5 Mammographic features

Several studies have described mammographic features of breast cancers after DBT and DM, however, as far as we know, no other prospective trial has compared mammographic features of recalled women screened with DBT versus DM.

Our study supports the finding of other studies confirming spiculated masses to be better visualized in DBT compared with DM [66; 96] and the association between spiculated masses and slow growing cancers, often luminal A-like [64; 118; 141]. However, spiculated mass was also the dominating group among non-luminal A-like cancers for women screened with DBT as well as DM.

Interestingly, there was similar proportion of luminal A-like cancers among invasive breast cancers after DBT compared with DM and the rates of luminal A-like cancers was also comparable for subsequent round of screen-detected breast cancers, in To-Be2 [132]. The results in the To-Be trials indicates that spiculated masses in DBT probably not only contribute to detection of small, low proliferation cancers, but also contribute to detect a substantial number of non-luminal A-like cancers. Similar results were reported also in the Malmö trial [120].

Indistinct mass is described as a suspicious finding for breast cancer in BI-RADS [95]. The results in the To-Be1-trial supports the statement which was even stronger for DBT versus DM in our study. Despite fewer recalled women due to indistinct masses for DBT, the same number of breast cancers presenting as indistinct masses was detected for DBT as for DM. A benign outcome after recall because of an indistinct mass was less frequent after DBT compared with DM, indicating that an indistinct mass is more likely to represent cancer in DBT than in DM.

Asymmetry was the most common feature among women recalled after screening with DBT and with DM. Only half as many women were recalled because of asymmetry after DBT (n=108) compared with DM (n=222). The likelihood of malignancy if recalled was rather low in both screening-modalities, 1.6% (7/444) for DBT and 2.1% (12/571) for DM in our study. The feature contributes with a rather low proportion of screen-detected cancers in DBT in To-Be1 and in several other studies [66; 113]. Consequently, it should be discussed whether fewer women should be recalled because of asymmetry if screened with DBT in the future.

Similarly, taking advantage of our results, radiologists might decide to refrain to recall women with circumscribed masses that are considered most likely benign, after taking into consideration factors like age, use of hormonal replacement treatment, density/location of the mass, and number of additional circumscribed masses looking like cysts.

In DM-screening fine linear branching calcifications are known to have a high association with malignancy compared with other calcification types. We did not find

statistical differences between DBT and DM for calcifications. Our study did not distinguish between fine linear branching calcifications and other types of calcifications in DBT and DM. This seems important to study in more detail in future studies. Further, there might be differences between radiologists regarding which features the individual radiologist perceive or decide to recall. Our study has not evaluated such differences.

Being aware of the increased likelihood of detecting slow-growing cancers among architectural distortions/spiculated masses/tissue abnormalities [42; 44; 117] brings up the discussion of screening and risk of overdiagnosis. When focusing on detecting rapidly growing, potentially killing cancers, without increasing false positives and overdiagnosis to undesirably high level, availability of prior mammograms could be considered important. We did not detect the same level of additional breast cancers classified as architectural distortions as other studies [66; 113]. The reason might partly be due to how prior mammograms, which were easily available up to 10 years back in time, were used.

In prospective European trials, different feature categories have been used, making comparison of results difficult [66; 96; 113]. No European studies have used all variables and categories described in BI-RADS Mammography Breast Imaging Lexicon [66; 96; 113]. As far as we have found, no studies have reported shape, all studies collapsed groups of margins. Several studies had some information regarding density. The reason for collapsing mammographic features into fewer categories in studies than in BI-RADS, could be the small numbers of breast cancers in each feature category, and therefore statistical challenges to achieve significant results in studies. Another reason might be related to whether the study includes benign and malignant lesions or only the latter.

The Oslo and the Malmö trial, collapsed circumscribed, obscured and indistinct masses into one group (mass with calcification was separated in the Oslo study). The Oslo Tomosynthesis Screening Trial used the following categories of features: circumscribed mass, spiculated mass, architectural distortion, asymmetric density,

calcifications and mass and calcifications. The Malmö trial reported five different features (spiculated mass, circumscribed mass, microcalcifications, architectural distortions and enlarged lymph nodes in addition to the category missing). If mass-categories of well-defined (circumscribed) often benign masses are collapsed into the same category as suspicious findings as indistinct (ill-defined) mass or microlobulated mass, results regarding mammographic features and screening-outcome might be difficult to interpret and to compare between studies.

In To-Be1, we used a slightly modified BI-RADS feature classification system; we did not use the term “special cases” but added the feature “mass with calcifications”. This feature had been used in the Oslo trial, further, Tabar had shown that breast cancers with the feature casting type calcifications (with or without in combination with a mass) have an unexpectedly poor prognosis [66; 116]. Thus, the feature seemed to be of special interest. However, some masses with calcifications would probably have been categorized as indistinct or spiculated masses if we had not used the category. In retrospect, perhaps a strict BI-RADS classification system would have been chosen.

Being aware of the better visualisation of at least some mammographic features in DBT compared with DM, and the differences between mammographic features in DBT and DM, it seems fair to have high expectations to results after DBT has been used in screening. The possibility of introducing AI in the screen reading of DBT may further boost diagnostic performance. However, more studies are needed before any conclusions can be drawn.

5.2 Other factors when implementing DBT

Studies published before start-up of ToBe1 mainly used equipment from other vendors (Hologic and Siemens) than GE. Being aware of the differences in DBT systems, studies performed with equipment from different vendors are needed. Caution should be used before extrapolating results from one DBT-system to another.

There is need for high-speed data lines between the screening unit and the PACS system and workstations. Further, the requirements regarding speed, workflow and storage capacity of the workstations are substantial when screen reading is performed with DBT.

When evaluating a potentially new screening tool, cost effectiveness should be considered. A study evaluating cost differences between DBT and DM regarding equipment, examination and reading time in To-Be, observed an increase of €8.5 per screened woman for DBT versus DM. The differences remained higher after recall assessment cost were added (€6.2). On the other hand, if DBT screening results lead to earlier detection of more breast cancers, fewer interval cancers and eventually a lower breast cancer mortality, this incremental cost may be acceptable. When a similar cost-effectiveness analysis is performed on the results after To-Be2, where recall rate of DBT were higher (3.9%) compared with To-Be1 (3.1%), and the detection rate was higher (8.6/1000 screened after DBT in To-Be2 versus 6.6/1000 in To-Be) the cost estimate will change. The estimated additional cost of DBT versus DM after To-Be1 and To-Be2 was €8.1. In a simulation, 500 deaths would be averted, and 2,300 life-years gained after ten rounds of DBT screening from 2018 inducing an additional screening cost of €29 million. Further, the incremental cost-effectiveness ratio indicated cost savings of €1,400 per life-year gained (from still unpublished results, Tron Moger). These results indicate that introducing DBT in BreastScreen Norway may be cost-effective. However, recall rates, sensitivity and specificity in screening programmes, treatment cost and willingness-to-pay levels should be considered before used as an argument for introducing routine screening with DBT in other countries.

5.3 Methodological considerations

5.3.1 Study design in To-Be1

Using a randomized design is state of the art, and optimal in a study setting like ours. All women attending screening in Bergen, embedded in BreastScreen Norway during 2016 and 2017, were invited to participate in the trial, and 89.3% of the women

consented and were thus included. The rather high rate of women accepting the invitation to participate in the study further strengthens the quality of the trial as the participants are considered representative of the screening population. The randomization of the participants ensured there were no systematic differences between the control group and the study group. The only difference between the DBT and DM groups in our trial was the screening technique. The fact that the radiologists could not be blinded for which screening method the radiologist were reading, should not be considered as a major limitation, if the radiologist aimed to read similar levels of DBT and DM mammograms, and as long as the results were not reported on a group level before the end of the study period.

5.3.2 Performance of the trial

Studies have shown significant differences between radiologists regarding time used in screen-reading, recall and detection rate in DBT and DM. Studies have also shown that several radiologists detect more cancers when reading DBT while others detect equal numbers when reading DM [66; 113]. Thus, the fact that we did not succeed in completely balancing the number of readings performed between radiologists, nor completely balance number of DBT and DM-readings for each radiologist is a limitation in our trial. If one or more of the radiologists reading substantial part of DM were “high quality DM-readers” with high detection rate after DM, this might potentially have influenced our results.

During the performance of To-Be1, the breast centre had no access to tomosynthesis guided or MRI-guided biopsy-device. It is well known that some few mammographic findings at DBT can be difficult or impossible to recognise using ultrasound or DM. An ultrasound or stereotactic guided biopsy could thus be very challenging to perform. Thus, it is possible that some few findings at DBT did not have a representative biopsy or were incorrectly dismissed at assessment. However, this is likely a minor limitation, as most findings are visible at ultrasound or DM.

5.3.3 Data quality

It has been mandatory by law for all hospitals to report cancer cases to the Cancer Registry of Norway since 1953 [24] and breast cancer reporting is close to 100% complete [3]. Further, a research assistant ensured the reporting, which also was merged with data from the Cancer Registry. Using an automated device to report breast density and radiation dose is considered a strength in this trial, reducing individual differences in perception of breast density and allowing a more accurate report of dose compared to machine reported dose.

No feedback regarding recall or detection rate was reported during the trial, which can represent a strength in our trial, as only findings on mammograms were considered, before radiologists decided whether to recall or not. If information on recall rates had been available for the radiologists, this could have introduced a trial effect.

The automated software reported a difference in median values of fibroglandular volume and mean volumetric breast density for women in the DBT versus the DM-arm. Further, proportions of women in the VDG 3 and VDG 4 groups were lower for DBT compared with DM (study 2). Differences in breast density are not expected in an RCT. However, such differences have been reported in other studies [111; 142], and can be explained by discrepancies in estimation of breast density by the software in DBT versus DM. The fact that the software was not able to define quite identical breast density categories in DBT and DM is a limitation.

Selection bias

Selection bias is a systematic error which occurs when the selection of individuals or data for analysis is performed in such a way that proper randomization is not achieved, thereby ensuring that the sample studied is not representative of the population intended to be analysed [143]. In To-Be1 the risk of selection bias seems minimal as the randomization to DBT or DM were done after the women had signed the agreement to participate and because of the high attendance rate. The randomization procedure was performed electronically at an individual level. There is

little reason to believe that the women or radiologists in Bergen are different compared with women or radiologists in other Norwegian cities. However, there might be a difference between radiologists in Norway and the US because of differences in screening programs and traditions.

We excluded women with implants, women with a history of breast cancer, and one woman with metastasis from melanoma. In study 2, additionally 121 women screened with DBT and 86 screened with DM were excluded because of random error causing lack of Volpara density results. With a study population of 14380 and 14369 these exclusion numbers seem too small to have largely influenced the results.

Confounding and mediators

Confounding refers to a situation in which other factors affects or confuses the outcome of a study [144]. Confounding has been present if a variable has affected both the dependant/outcome variable and the independent/exposure variable, and a false association thus seem to be present. In research, the dependant variable is the variable that can be considered being "influenced" by other variables, for example breast cancer, in this study. The independent variable is the variable that is "influencing" the dependant variable, e.g. DBT versus DM. We aimed to explore the association between screening technique (independent/exposure variable) and early performance indicators (e.g. reading time, radiation dose, recall; the dependent/outcome variables). Because of the randomization procedure in this RCT, confounding should not be a major challenge [145]. There should not be any differences between the independent variables (women exposed to DBT versus DM).

We have discussed if the radiologists experiences in screen-reading DBT might have influenced the number of cases discussed in consensus, recalled and cancers detected (more experienced radiologists may have led to higher recall- and breast cancer detection rates in To-Be2 versus To-Be1 [132]). However, experience is still not considered a confounding factor; it did not affect the independent variable (DBT or DM) in our study. It may, however, have affected the dependent factor (recall, cancer

detection and other) among the DBT-screened, thus, it may represent a potential mediator.

The eight participating radiologists did not read the same number of images in To-Be1, which was not optimal. However, another, and probably even more important limitation is the fact that two of the three radiologists with highest volume of read DBT-images, read only half the number, or almost the double volume of DM-images during the first year (Table 5 in Appendix Study 1). Studies have shown there might be considerable differences between and within radiologists when it comes to recall and detection rates in DBT and DM [66; 113]. This is not a confounding factor, but might be considered a mediator.

The lack of DBT guided biopsy may also have influenced outcome on an individual basis, as some more breast cancers could have been detected among women recalled in DBT if the equipment had been available. It is, however, unlikely that the study outcome would have significantly differed, as the number of breast cancers which were not possible to detect or biopsy at assessment with other equipment, were probably very few. Thus, the lack of DBT guided biopsy is not considered a confounder, but might be yet another mediator.

The To-Be-trials have several strengths. The assessment of recalled women were performed by the same group of radiologists that performed the screen reading. All radiologists randomly participated in the assessment of women recalled after DBT and after DM. At assessment, the same equipment was available for all women; DM, DBT, ultrasound, biopsies (ultrasound or stereotactic guided), eventually MRI. Blinding of the radiologists was impossible. However, all women invited to screening at the screening centre in Bergen were invited in the study; some women lived in the city and other women lived in surrounding areas. This supports the idea of a random selection regarding exposure and a lower risk of confounding biases.

Internal validity

When evaluating a study, methodological considerations regarding internal and external validity should be considered. Internal validity is the extent to which we can

be confident that the cause-and-effect relationship established in a study cannot be explained by other factors [146; 147]. Internal validity expresses that the results are correct and valid for the studied sample. RCTs often have high internal validity due to the randomization procedure, and because of the defined inclusion and exclusion criteria [148]. We consider the internal validity in our RCT to be high.

External validity and generalisation

The external validity relates to how applicable the findings of a study are in the real world. It indicates the extent to which the results are valid under other conditions and for other samples, i.e. it relates to generalizability [146; 147]. The fact that To-Be1 is a single centre trial is a limitation related to external validity. Radiologists are human beings, interpretation of mammograms is a subjective task. The fact that few radiologists at a single centre participated in the screen reading, and the fact that some of the participating radiologists read larger volumes of the mammograms, might be limitations to generalizability in To-Be1.

Further, the very modest experience with DBT in screening among radiologists before start-up of the trial might hamper external validity in To-Be1: results on recall rate, detection rate, reading time and other early performance measures changed during the study periods of the To-Be trials, even though the radiologists were the same. A heavy work load on participating radiologists might also have affected results. These factors may have slightly hampered the generalizability of the results.

Still, there are factors we consider can be generalized and applied to a broader context. Our findings from ToBe1 and later published studies from To-Be2 indicate that performing population based screening with DBT is possible. Technical equipment can manage the huge increase in workload when 4 images in DM is replaced with approximately 200-250 images per woman in a DBT, and the storage capacity increases from ~30 MB in DM to ~1000-3000 MB in DBT for each woman. Radiologists spend more time reading DBT compared with DM. Screening with DBT versus DM increases costs of equipment, examination and reading [81]. Thus, screening with DBT is more resource consuming, but possible to perform. Results

from study 1 and study 2 indicate that DBT screening may represent a slightly more precise technique; with lower consensus- and recall rate, and lower rate of false positives in women with low breast density. Furthermore, our studies indicate that screening with DBT is safe; in spite of lower recall rate after DBT in our trial; rate of detected breast cancers was at the same level compared with DM for all density categories.

Distribution of mammographic features was different for DBT compared with DM. Our findings in study 3 indicate there is potential for further increase in benefit/harm ratio of screening. Radiologists might be encouraged to consider dismissing less suspicious asymmetries and some circumscribed masses after DBT. Calcifications should be carefully evaluated, and more attention should be addressed on indistinct masses, spiculated masses and architectural distortions, using information from priors to ensure that reasonable decisions are made.

Factors related to equipment, i.e. angle on tomosynthesis, reconstruction algorithm for DBT and the quality of the synthetic 2D image, and the rather low radiation dose used, might also have influenced the outcome. Thus, caution should be used before extrapolating results from the To-Be1 trial to breast centres using equipment from other vendors.

5.4 Conclusion

The results shown in ToBe1 indicated that screening with DBT was as least as good as screening with DM with respect to recalls, false positives and screen-detected cancer, and thus safe for the women. After one year using DBT and DM in an RCT we found that radiologists spent more time reading DBT compared with DM. Mean glandular dose did not differ either for DBT versus DM or breast density. Further, we found recall- and consensus-rates to be lower in DBT versus DM in To-Be1, while the rate of screen-detected cancer did not differ statistically.

When exploring differences in early performance measures by automated breast density in To-Be1, we found DBT to be superior to DM in women with lower breast

density; similar cancer detection, lower recall and less false positive screening results in the two lowest density categories (69% of the women). Adjusted relative risk of recall, false positive screening result and screen-detected cancers increased with breast density (VDG1-4) in DBT, whereas no statistically differences were observed for DM.

More women were recalled and diagnosed with breast cancer because of spiculated mass after DBT compared with DM. Spiculated mass was the dominant feature of breast cancer in women screened with DBT, while calcifications was the most frequent feature in women screened with DM. Asymmetry, indistinct and obscured mass was less frequent in women with a false positive screening result (recalled, but no breast cancer detected) after screening with DBT versus DM. Analysing results in our study might encourage radiologists to more strictly consider which features to recall in DBT screening. Less attention might be given to some asymmetries and some circumscribed masses, and more attention to features where there is higher susceptibility of cancers. More knowledge of the differences in distribution of mammographic features and their association with screening outcome might contribute to further improve the benefits of DBT as a screening tool for breast cancer.

Screening with DBT is more resource consuming than screening with DM according to time spent on screen reading and consensus.

6. FUTURE PERSPECTIVES

When bearing in mind that DBT consists of ~200-250 images per women, compared to 4 images in a DM, it seems obvious that humans might miss important findings when performing batch-reading of several DBT-examinations. The recommended volume in the batches of DBT examinations should thus be lower than for DM. Introducing AI in screening could reduce the work load for the radiologists and probably the risk of missing important findings [149-151]. Studies with DBT and AI, both in retrospective and prospective screening cohorts, are thus warranted. However, in parallel, training and developments of the algorithms are needed.

The results indicate limited sensitivity for DBT for women with high mammographic density. Running the same analysis on mammographic features for breast cancer cases detected in To-Be2 might be of interest to strengthen the findings in To-Be1. Further studies exploring optimal supplemental screening techniques in women with densest breasts are warranted.

Mammographic features in DBT and their association with screening outcome should be studied in more detail, aiming an increased benefit and decreased harms of screening. Results from our studies should encourage particular focus on indistinct and spiculated masses, calcifications and architectural distortions when reading DBT. Hopefully, the use of AI-guided tools in the future may provide diagnostic support for radiologists reading mammograms that will yield higher sensitivity for detecting clinically relevant cancers [152; 153] and reduce the number of missed and slow growing cancers in screening examinations [150; 153]).

Although DBT is a safe screening method for women, costs and long-term health benefits should be further explored to assess whether DBT should replace DM in future breast cancer screening programs.

7. REFERENCES

1 DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A (2015) International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev 24:1495-1506

2 Cancer Registry of Norway (2020) Cancer in Norway 2019 - Cancer incidence, mortality, survival and prevalence in Norway, Accessed 2021-03-29. Available via https://www.kreftregisteret.no/globalassets/cancer-in-norway/2019/cin_report.pdf

3 Cancer Registry of Norway (2021) Cancer in Norway 2020 - Cancer incidence, mortality, survival and prevalence in Norway, Accessed 2022-05-25 Available via https://www.kreftregisteret.no/globalassets/cancer-in-norway/2020/cin-2020.pdf

4 Årsrapport 2020 med resultater og forbedringstiltak fra Nasjonalt kvalitetsregister for brystkreft. Accessed 2022-05-25. (2021) Oslo. Available via https://www.kvalitetsregistre.no/sites/default/files/2021-10/%C3%85rsrapport%20Nasjonalt%20kvalitetsregister%20for%20brystkreft%202020.pdf

5 Dibden A, Offman J, Duffy SW, Gabe R (2020) Worldwide Review and Meta-Analysis of Cohort Studies Measuring the Effect of Mammography Screening Programmes on Incidence-Based Breast Cancer Mortality. Cancers (Basel) 12

6 Duffy SW, Tabár L, Yen AM et al (2021) Beneficial Effect of Consecutive Screening Mammography Examinations on Mortality from Breast Cancer: A Prospective Study. Radiology. 10.1148/radiol.2021203935:203935

7 Sebuodegard S, Botteri E, Hofvind S (2019) Breast cancer mortality after implementation of organized population-based breast cancer screening in Norway. J Natl Cancer Inst. 10.1093/jnci/djz220

8 Duffy SW, Tabár L, Yen AM et al (2020) Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women. Cancer 126:2971-2979

9 Sung H, Ferlay J, Siegel R et al (2020) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. Accessed 2021-08-05 International Agency for Research on Cancer (IARC) World Health Organization. Available via https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf. 10.3322/caac.21660.

10 McPherson K, Steel CM, Dixon JM (2000) ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. Bmj 321:624-628

11 Boyd N, Martin L, Stone J, Little L, Minkin S, Yaffe M (2002) A longitudinal study of the effects of menopause on mammographic features. Cancer Epidemiol Biomarkers Prev 11:1048-1053

12 Boyd NF, Byng JW, Jong RA et al (1995) Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study. J Natl Cancer Inst 87:670-675

13 Brand JS, Czene K, Eriksson L et al (2013) Influence of lifestyle factors on mammographic density in postmenopausal women. PLoS One 8:e81876

14 Byrne C, Schairer C, Wolfe J et al (1995) Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 87:1622-1629
15 McCormack VA, dos Santos Silva I (2006) Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 15:1159-1169
16 Roman M, Graff-Iversen S, Weiderpass E et al (2016) Postmenopausal Hormone Therapy and Breast Cancer Prognostic Characteristics: A Linkage between Nationwide Registries. Cancer Epidemiol Biomarkers Prev 25:1464-1473
17 Strand F, Humphreys K, Holm J et al (2018) Long-term prognostic implications of risk factors associated with tumor size: a case study of women regularly attending screening. Breast Cancer Res 20:31
18 Castells X, Torá-Rocamora I, Posso M et al (2016) Risk of Breast Cancer in Women with False-Positive Results according to Mammographic Features. Radiology 280:379-386
19 Hofvind S, Sagstad S, Sebuødegård S, Chen Y, Roman M, Lee CI (2018) Interval Breast Cancer Rates and Histopathologic Tumor Characteristics after False-Positive Findings at Mammography in a Population-based Screening Program. Radiology 287:58-67
20 von Euler-Chelpin M, Risør LM, Thorsted BL, Vejborg I (2012) Risk of breast cancer after false-positive test results in screening mammography. J Natl Cancer Inst 104:682-689
21 Naume B, Schlichting E, Eikesdal HP, Hagen AI, Blix ES (2021) National recommendations of Diagnostics and Treatment of Breast Cancer. Norwegian Directorate of Health, Oslo. Available via https://www.helsedirektoratet.no/retningslinjer/brystkreft-handlingsprogram. Accessed 2021-03-01
22 Cardoso F, Kyriakides S, Ohno S et al (2019) Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. Ann Oncol 30:1194-1220
23 Nass SJ, Henderson IC, Lashof JC (2001) Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer. National Academies Press (US) the National Academy of Sciences. , Washington (DC). Available via https://www.ncbi.nlm.nih.gov/books/NBK222348/. 10.17226/10030
24 Hofvind S, Tsuruda K, Mangerud G, Ertzaas AK, et al (2017) The Norwegian Breast Cancer Screening Program, 1996-2016: Celebrating 20 years of organised mammographic screening. Accessed 2021-01-05. Cancer Registry of Norway, Oslo. Available via https://www.kreftregisteret.no/globalassets/cancer-in-norway/2016/mammo_cin2016_special_issue_web.pdf
25 Skaane P (2009) Studies comparing screen-film mammography and full-field digital mammography in breast cancer screening: updated review. Acta Radiol 50:3-14

26 Ertzaas AH, M; Hantho, C; Hammond, RL; Larsen, AN; Music, J; , Westermann LT, L; Wåde, GG; Hofvind, S (2021) Quality manual for radiography in BreastScreen Norway, Accessed 2021-12-16. The Cancer Registry of Norway, Oslo. Available via https://www.kreftregisteret.no/globalassets/publikasjoner-og-rapporter/mammografiprogrammet/2021-delmanual-for-radiografi.pdf

27 Freer PE (2015) Mammographic breast density: impact on breast cancer risk and implications for screening. Radiographics 35:302-315

28 Pisano ED, Hendrick RE, Yaffe MJ et al (2008) Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology 246:376-383

29 Dobbins JT, 3rd, Godfrey DJ (2003) Digital x-ray tomosynthesis: current state of the art and clinical potential. Phys Med Biol 48:R65-106

30 Gur D (2007) Tomosynthesis: potential clinical role in breast imaging. AJR Am J Roentgenol 189:614-615

31 Houssami N (2018) Evidence on Synthesized Two-dimensional Mammography Versus Digital Mammography When Using Tomosynthesis (Three-dimensional Mammography) for Population Breast Cancer Screening. Clin Breast Cancer 18:255-260.e251

32 Holen Å S (2016) Illustration of tomosynthesis. Cancer Registry of Norway. Available via https://www.kreftregisteret.no/en/Research/Projects/to-be-studies/

33 Gilbert FJ, Tucker L, Young KC (2016) Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol 71:141-150

34 Michell MJ, Batohi B (2018) Role of tomosynthesis in breast imaging going forward. Clin Radiol 73:358-371

35 Friedewald SM, Bhole S, Wang L, Gupta D (2019) Digital Breast Tomosynthesis: Clinical Operations. Journal of Breast Imaging 1:122-126

36 Lai Y-C, Ray KM, Mainprize JG, Kelil T, Joe BN (2020) Digital Breast Tomosynthesis: Technique and Common Artifacts. Journal of Breast Imaging 2:615-628

37 Ikejimba LC, Salad J, Graff CG et al (2021) Assessment of task-based performance from five clinical DBT systems using an anthropomorphic breast phantom. Medical Physics 48:1026-1038

38 Giuliano AE, Connolly JL, Edge SB et al (2017) Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. CA Cancer J Clin 67:290-303

39 Silverstein MJ, Poller DN, Waisman JR et al (1995) Prognostic classification of breast ductal carcinoma-in-situ. Lancet 345:1154-1157

40 Curigliano G, Burstein HJ, Winer EP et al (2018) 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. Ann Oncol 29:2153

41 Goldhirsch A, Winer EP, Coates AS et al (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International

Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 24:2206-2223
42 Boisserie-Lacroix M, Hurtevent-Labrot G, Ferron S, Lippa N, Bonnefoi H, Mac Grogan G (2013) Correlation between imaging and molecular classification of breast cancers. Diagn Interv Imaging 94:1069-1080
43 Hofvind S, Hovda T, Holen AS et al (2018) Digital Breast Tomosynthesis and Synthetic 2D Mammography versus Digital Mammography: Evaluation in a Population-based Screening Program. Radiology 287:787-794
44 Sturesdotter L, Sandsveden M, Johnson K, Larsson AM, Zackrisson S, Sartor H (2020) Mammographic tumour appearance is related to clinicopathological factors and surrogate molecular breast cancer subtype. Sci Rep 10:20814
45 Niell BL, Freer PE, Weinfurtner RJ, Arleo EK, Drukteinis JS (2017) Screening for Breast Cancer. Radiol Clin North Am 55:1145-1162
46 Wilson JMG, Junger G (1968) Principles and practice of screening for disease. World Health Organization. Available via https://apps.who.int/iris/handle/10665/37650
47 Norwegian Directorate of Health (2014) Styringsstruktur og strategi for nasjonale screeningprogrammer, Accessed 2022-05-25. Norwegian Directorate of Health Oslo. Available via https://www.legeforeningen.no/contentassets/9508a2d19cd2477c8371ba939943d042/hoeringsutkast-fra-helsedirektoratet-styringsstruktur-og-strategi-for-nasjonale-screeningprogrammer.pdf
48 European Commission Initiative On Breast Cancer (2021) European guidelines on breast cancer screening and diagnosis. Accessed 2021-06-30. European Commission. Available via https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/screening-ages-and-frequencies
49 Lauby-Secretan B, Scoccianti C, Loomis D et al (2015) Breast-cancer screening--viewpoint of the IARC Working Group. N Engl J Med 372:2353-2358
50 Peintinger F (2019) National Breast Screening Programs across Europe. Breast Care (Basel) 14:354-358
51 Bento MJ, Gonçalves G, Aguiar A, Castro C, Veloso V, Rodrigues V (2015) Performance indicators evaluation of the population-based breast cancer screening programme in Northern Portugal using the European Guidelines. Cancer Epidemiol 39:783-789
52 Broeders MJ, Scharpantgen A, Ascunce N et al (2005) Comparison of early performance indicators for screening projects within the European Breast Cancer Network: 1989-2000. Eur J Cancer Prev 14:107-116
53 Alabousi M, Wadera A, Kashif Al-Ghita M et al (2021) Performance of Digital Breast Tomosynthesis, Synthetic Mammography and Digital Mammography in Breast Cancer Screening: A Systematic Review and Meta-Analysis. J Natl Cancer Inst 113:680-690
54 Marinovich ML, Hunter KE, Macaskill P, Houssami N (2018) Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. J Natl Cancer Inst 110:942-949

55 Heindel W, Weigel S, Gerß J et al (2022) Digital breast tomosynthesis plus synthesised mammography versus digital screening mammography for the detection of invasive breast cancer (TOSYMA): a multicentre, open-label, randomised, controlled, superiority trial. Lancet Oncol 23:601-611

56 Song JW, Chung KC (2010) Observational studies: cohort and case-control studies. Plast Reconstr Surg 126:2234-2242

57 Boyd NF, Guo H, Martin LJ et al (2007) Mammographic density and the risk and detection of breast cancer. N Engl J Med 356:227-236

58 Strand F, Azavedo E, Hellgren R et al (2019) Localized mammographic density is associated with interval cancer and large breast cancer: a nested case-control study. Breast Cancer Res 21:8

59 Hofvind S, Holen Å, Román M, Sebuødegård S, Puig-Vives M, Akslen L (2016) Mode of detection: an independent prognostic factor for women with breast cancer. J Med Screen 23:89-97

60 Houssami N, Hunter K (2017) The epidemiology, radiology and biological characteristics of interval breast cancers in population mammography screening. NPJ Breast Cancer 3:12

61 Hovda T, Hoff SR, Larsen M, Romundstad L, Sahlberg KK, Hofvind S (2022) True and Missed Interval Cancer in Organized Mammographic Screening: A Retrospective Review Study of Diagnostic and Prior Screening Mammograms. Acad Radiol 29,Supplement 1:S180-S191

62 Bernardi D, Macaskill P, Pellegrini M et al (2016) 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. Lancet Oncol 17:1105-1113

63 Ciatto S, Houssami N, Bernardi D et al (2013) Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 14:583-589

64 Lang K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S (2016) 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. Eur Radiol 26:184-190

65 Skaane P, Bandos AI, Gullien R et al (2013) Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. Eur Radiol 23:2061-2071

66 Skaane P, Bandos AI, Gullien R et al (2013) Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 267:47-56

67 Houssami N, Zackrisson S, Blazek K et al (2021) Meta-analysis of prospective studies evaluating breast cancer detection and interval cancer rates for digital breast tomosynthesis versus mammography population screening. Eur J Cancer 148:14-23

68 Chong A, Weinstein SP, McDonald ES, Conant EF (2019) Digital Breast Tomosynthesis: Concepts and Clinical Practice. Radiology 292:1-14

69 Giampietro RR, Cabral MVG, Lima SAM, Weber SAT, Dos Santos Nunes-Nogueira V (2020) Accuracy and Effectiveness of Mammography versus Mammography and Tomosynthesis for Population-Based Breast Cancer Screening: A Systematic Review and Meta-Analysis. Sci Rep 10:7991
70 McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF (2016) Effectiveness of Digital Breast Tomosynthesis Compared With Digital Mammography: Outcomes Analysis From 3 Years of Breast Cancer Screening. JAMA Oncol 2:737-743
71 Skaane P, Sebuødegård S, Bandos AI et al (2018) Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. Breast Cancer Res Treat 169:489-496
72 Johnson K, Lång K, Ikeda DM, Åkesson A, Andersson I, Zackrisson S (2021) Interval Breast Cancer Rates and Tumor Characteristics in the Prospective Population-based Malmö Breast Tomosynthesis Screening Trial. Radiology 299:559-567
73 Houssami N (2017) Overdiagnosis of breast cancer in population screening: does it make breast screening worthless? Cancer Biol Med 14:1-8
74 (2012) The benefits and harms of breast cancer screening: an independent review. Lancet 380:1778-1786
75 Lynge E, Beau AB, von Euler-Chelpin M et al (2020) Breast cancer mortality and overdiagnosis after implementation of population-based screening in Denmark. Breast Cancer Res Treat 184:891-899
76 Dang PA, Freer PE, Humphrey KL, Halpern EF, Rafferty EA (2014) Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations. Radiology 270:49-56
77 Zuckerman SP, Conant EF, Keller BM et al (2016) Implementation of Synthesized Two-dimensional Mammography in a Population-based Digital Breast Tomosynthesis Screening Program. Radiology 281:730-736
78 Maldera A, De Marco P, Colombo PE, Origgi D, Torresin A (2017) Digital breast tomosynthesis: Dose and image quality assessment. Phys Med 33:56-67
79 Gennaro G, Bernardi D, Houssami N (2018) Radiation dose with digital breast tomosynthesis compared to digital mammography: per-view analysis. Eur Radiol 28:573-581
80 Svahn TM, Houssami N, Sechopoulos I, Mattsson S (2015) Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. Breast 24:93-99
81 Moger TA, Swanson JO, Holen Å S, Hanestad B, Hofvind S (2019) Cost differences between digital tomosynthesis and standard digital mammography in a breast cancer screening programme: results from the To-Be trial in Norway. Eur J Health Econ 20:1261-1269
82 Nelson HD, Pappas M, Cantor A, Griffin J, Daeges M, Humphrey L (2016) Harms of Breast Cancer Screening: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation. Ann Intern Med 164:256-267

83 Kopans DB (2008) Basic physics and doubts about relationship between mammographically determined tissue density and breast cancer risk. Radiology 246:348-353

84 Checka CM, Chun JE, Schnabel FR, Lee J, Toth H (2012) The relationship of mammographic density and age: implications for breast cancer screening. AJR Am J Roentgenol 198:W292-295

85 Maskarinec G, Pagano I, Lurie G, Kolonel LN (2006) A longitudinal investigation of mammographic density: the multiethnic cohort. Cancer Epidemiol Biomarkers Prev 15:732-739

86 Moshina N, Sebuodegard S, Lee CI et al (2018) Automated Volumetric Analysis of Mammographic Density in a Screening Setting: Worse Outcomes for Women with Dense Breasts. Radiology 288:343-352

87 Nazari SS, Mukherjee P (2018) An overview of mammographic density and its association with breast cancer. Breast Cancer 25:259-267

88 Boyd N, Martin L, Gunasekara A et al (2009) Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes. Cancer Epidemiol Biomarkers Prev 18:1754-1762

89 Duffy SW, Morrish OWE, Allgood PC et al (2018) Mammographic density and breast cancer risk in breast screening assessment cases and women with a family history of breast cancer. Eur J Cancer 88:48-56

90 Vachon CM, van Gils CH, Sellers TA et al (2007) Mammographic density, breast cancer risk and risk prediction. Breast Cancer Res 9:217

91 Chiu SY, Duffy S, Yen AM, Tabár L, Smith RA, Chen HH (2010) Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. Cancer Epidemiol Biomarkers Prev 19:1219-1228

92 Baglietto L, Krishnan K, Stone J et al (2014) Associations of mammographic dense and nondense areas and body mass index with risk of breast cancer. Am J Epidemiol 179:475-483

93 Byrne C, Ursin G, Martin CF et al (2017) Mammographic Density Change With Estrogen and Progestin Therapy and Breast Cancer Risk. J Natl Cancer Inst 109

94 Conner P, Svane G, Azavedo E et al (2004) Mammographic breast density, hormones, and growth factors during continuous combined hormone therapy. Fertil Steril 81:1617-1623

95 D'Orsi CJ, Sicles EA, Mendelson EB, Morris EA, et al (2013) ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. American College of Radiology, Reston, VA. Available via https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

96 Zackrisson S, Lang K, Rosso A et al (2018) One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol 19:1493-1503

97 Moshina N, Roman M, Sebuødegård S, Waade GG, Ursin G, Hofvind S (2018) Comparison of subjective and fully automated methods for measuring mammographic density. Acta Radiol 59:154-160

98 Sprague BL, Conant EF, Onega T et al (2016) Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study. Ann Intern Med 165:457-464
99 Astley SM, Harkness EF, Sergeant JC et al (2018) A comparison of five methods of measuring mammographic density: a case-control study. Breast Cancer Res 20:10
100 Li J, Szekely L, Eriksson L et al (2012) High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. Breast Cancer Res 14:R114
101 van der Waal D, den Heeten GJ, Pijnappel RM et al (2015) Comparing Visually Assessed BI-RADS Breast Density and Automated Volumetric Breast Density Software: A Cross-Sectional Study in a Breast Cancer Screening Setting. PLoS One 10:e0136667
102 Eng A, Gallant Z, Shepherd J et al (2014) Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. Breast Cancer Res 16:439
103 Singh JM, Fallenberg EM, Diekmann F et al (2013) Volumetric breast density assessment: reproducibility in serial examinations and comparison with visual assessment. Rofo 185:844-848
104 Volpara Health Technologies. VolparaSolutions—VolparaDensity. Accessed 2021-09-21. Available via https://www.volparahealth.com/breast-health-platform/products/scorecard/
105 Machida Y, Saita A, Namba H, Fukuma E (2016) Automated volumetric breast density estimation out of digital breast tomosynthesis data: feasibility study of a new software version. Springerplus 5:780
106 Pisano ED, Gatsonis C, Hendrick E et al (2005) Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 353:1773-1783
107 Wanders JO, Holland K, Veldhuis WB et al (2017) Volumetric breast density affects performance of digital screening mammography. Breast Cancer Res Treat 162:95-103
108 Bjelland EK, Hofvind S, Byberg L, Eskild A (2018) The relation of age at menarche with age at natural menopause: a population study of 336 788 women in Norway. Hum Reprod 33:1149-1157
109 Bertrand KA, Tamimi RM, Scott CG et al (2013) Mammographic density and risk of breast cancer by age and tumor characteristics. Breast Cancer Res 15:R104
110 Moshina N, Ursin G, Hoff SR et al (2015) Mammographic density and histopathologic characteristics of screen-detected tumors in the Norwegian Breast Cancer Screening Program. Acta Radiol Open. Doi:10.1177/2058460115604340
111 Osteras BH, Martinsen ACT, Gullien R, Skaane P (2019) Digital Mammography versus Breast Tomosynthesis: Impact of Breast Density on Diagnostic Performance in Population-based Screening. Radiology 293:60-68

112 Conant EF, Barlow WE, Herschorn SD et al (2019) Association of Digital Breast Tomosynthesis vs Digital Mammography With Cancer Detection and Recall Rates by Age and Breast Density. JAMA Oncol 5:635-642

113 Romero Martín S, Raya Povedano JL, Cara García M, Santos Romero AL, Pedrosa Garriguet M, Álvarez Benito M (2018) Prospective study aiming to compare 2D mammography and tomosynthesis + synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis. Eur Radiol 28:2484-2491

114 Bernardi D, Ciatto S, Pellegrini M et al (2012) Prospective study of breast tomosynthesis as a triage to assessment in screening. Breast Cancer Res Treat 133:267-271

115 Gilbert FJ, Tucker L, Gillan MG et al (2015) The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme--a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. Health Technol Assess 19:i-xxv, 1-136

116 Tabar L, Tony Chen HH, Amy Yen MF et al (2004) Mammographic tumor features can predict long-term outcomes reliably in women with 1-14-mm invasive breast carcinoma. Cancer 101:1745-1759

117 Taneja S, Evans AJ, Rakha EA, Green AR, Ball G, Ellis IO (2008) The mammographic correlations of a new immunohistochemical classification of invasive breast cancer. Clin Radiol 63:1228-1235

118 Evans AJ, Pinder SE, James JJ, Ellis IO, Cornford E (2006) Is mammographic spiculation an independent, good prognostic factor in screening-detected invasive breast cancer? AJR Am J Roentgenol 187:1377-1380

119 Lamb PM, Perry NM, Vinnicombe SJ, Wells CA (2000) Correlation between ultrasound characteristics, mammographic findings and histological grade in patients with invasive ductal carcinoma of the breast. Clin Radiol 55:40-44

120 Johnson K, Zackrisson S, Rosso A et al (2019) Tumor Characteristics and Molecular Subtypes in Breast Cancer Screening with Digital Breast Tomosynthesis: The Malmo Breast Tomosynthesis Screening Trial. Radiology 293:273-281

121 Mariscotti G, Durando M, Houssami N et al (2016) Digital breast tomosynthesis as an adjunct to digital mammography for detecting and characterising invasive lobular cancers: a multi-reader study. Clin Radiol 71:889-895

122 Ray KM, Turner E, Sickles EA, Joe BN (2015) Suspicious Findings at Digital Breast Tomosynthesis Occult to Conventional Digital Mammography: Imaging Features and Pathology Findings. Breast J 21:538-542

123 Kim JY, Kang HJ, Shin JK et al (2017) Biologic Profiles of Invasive Breast Cancers Detected Only With Digital Breast Tomosynthesis. AJR Am J Roentgenol 209:1411-1418

124 Skaane P (2017) Breast cancer screening with digital breast tomosynthesis. Breast Cancer 24:32-41

125 Lilleborge M, Falk RS, Hovda T, Holmen MM, Ursin G, Hofvind S (2022) Patterns of aggressiveness: risk of progression to invasive breast cancer by mammographic features of calcifications in screen-detected ductal carcinoma in situ. Acta Radiol 63:586-595
126 Tagliafico A, Mariscotti G, Durando M et al (2015) Characterisation of microcalcification clusters on 2D digital mammography (FFDM) and digital breast tomosynthesis (DBT): does DBT underestimate microcalcification clusters? Results of a multicentre study. Eur Radiol 25:9-14
127 Hofvind S, Holen Å S, Aase HS et al (2019) Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. Lancet Oncol 20:795-805
128 Hudson S, Vik Hjerkind K, Vinnicombe S et al (2018) Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk. Breast Cancer Res 20:156
129 Ronckers CM, Erdmann CA, Land CE (2005) Radiation and breast cancer: a review of current evidence. Breast Cancer Res 7:21-32
130 Moshina N, Roman M, Waade GG, Sebuødegård S, Ursin G, Hofvind S (2018) Breast compression parameters and mammographic density in the Norwegian Breast Cancer Screening Programme. Eur Radiol 28:1662-1672
131 Pattacini P, Nitrosi A, Giorgi Rossi P et al (2018) Digital Mammography versus Digital Mammography Plus Tomosynthesis for Breast Cancer Screening: The Reggio Emilia Tomosynthesis Randomized Trial. Radiology 288:375-385
132 Hofvind S, Moshina N, Holen Å S et al (2021) Interval and Subsequent Round Breast Cancer in a Randomized Controlled Trial Comparing Digital Breast Tomosynthesis and Digital Mammography Screening. Radiology 300:66-76
133 Taourel P (2021) Interval Breast Cancer after Digital Breast Tomosynthesis-based Screening: A Glimmer of Hope. Radiology 300:77-78
134 Prospective Trial of Digital Breast Tomosynthesis (DBT) in Breast Cancer Screening. Accessed 2022-05-11. https://ClinicalTrials.gov/show/NCT03733106
135 Breast Cancer Screening: Digital Breast Tomosynthesis Versus Digital 2D Mammography. Accessed 2022-05-11. https://ClinicalTrials.gov/show/NCT03377036
136 Digital Tomosynthesis Mammography and Digital Mammography in Screening Patients for Breast Cancer.Accessed 2022-05-11. https://ClinicalTrials.gov/show/NCT03233191
137 Burnside ES, Vulkan D, Blanks RG, Duffy SW (2018) Association between Screening Mammography Recall Rate and Interval Cancers in the UK Breast Cancer Service Screening Program: A Cohort Study. Radiology 288:47-54
138 Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R (2001) Association of recall rates with sensitivity and positive predictive values of screening mammography. AJR Am J Roentgenol 177:543-549
139 Mann RM, Athanasiou A, Baltzer PAT et al (2022) Breast cancer screening in women with extremely dense breasts recommendations of the European

Society of Breast Imaging (EUSOBI). Eur Radiol. 10.1007/s00330-022-08617-6
140 Paci E (2012) Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. J Med Screen 19 Suppl 1:5-13
141 Liu S, Wu XD, Xu WJ, Lin Q, Liu XJ, Li Y (2016) Is There a Correlation between the Presence of a Spiculated Mass on Mammogram and Luminal A Subtype Breast Cancer? Korean J Radiol 17:846-852
142 Gastounioti A, McCarthy AM, Pantalone L, Synnestvedt M, Kontos D, Conant EF (2019) Effect of Mammographic Screening Modality on Breast Density Assessment: Digital Mammography versus Digital Breast Tomosynthesis. Radiology 291:320-327
143 Vetter TR, Mascha EJ (2017) Bias, Confounding, and Interaction: Lions and Tigers, and Bears, Oh My! Anesth Analg 125:1042-1048
144 Grimes DA, Schulz KF (2002) Bias and causal associations in observational research. Lancet 359:248-252
145 Guyatt GH, Oxman AD, Vist G et al (2011) GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol 64:407-415
146 Bolarinwa OA (2015) Principles and methods of validity and reliability testing of questionnaires used in social and health science researches. Niger Postgrad Med J 22:195-201
147 Pripp AH (2018) Validitet. Tidsskr Nor Laegeforen 138:1264
148 Spieth PM, Kubasch AS, Penzlin AI, Illigens BM, Barlinn K, Siepmann T (2016) Randomized controlled trials - a matter of design. Neuropsychiatr Dis Treat 12:1341-1349
149 Lång K, Dustler M, Dahlblom V, Åkesson A, Andersson I, Zackrisson S (2021) Identifying normal mammograms in a large screening population using artificial intelligence. Eur Radiol 31:1687-1692
150 Lång K, Hofvind S, Rodríguez-Ruiz A, Andersson I (2021) Can artificial intelligence reduce the interval cancer rate in mammography screening? Eur Radiol 31:5940-5947
151 Raya-Povedano JL, Romero-Martín S, Elías-Cabot E, Gubern-Mérida A, Rodríguez-Ruiz A, Álvarez-Benito M (2021) AI-based Strategies to Reduce Workload in Breast Cancer Screening with Mammography and Tomosynthesis: A Retrospective Evaluation. Radiology 300:57-65
152 Graewingholt A, Duffy S (2021) Retrospective comparison between single reading plus an artificial intelligence algorithm and two-view digital tomosynthesis with double reading in breast screening. J Med Screen 28:365-368
153 Kim HE, Kim HH, Han BK et al (2020) Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digit Health 2:e138-e148

8. Articles 1-3 with supplemental material

Paper 1

European Radiology (2019) 29:1175–1186
https://doi.org/10.1007/s00330-018-5690-x

BREAST

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

Hildegunn S. Aase[1,2] · Åsne S Holen[3] · Kristin Pedersen[3] · Nehmat Houssami[4] · Ingfrid S. Haldorsen[1,2] · Sofie Sebuødegård[3] · Berit Hanestad[1] · Solveig Hofvind[3,5]

Received: 1 May 2018 / Revised: 19 July 2018 / Accepted: 31 July 2018 / Published online: 29 August 2018

Abstract

Objectives To describe a randomized controlled trial (RCT) of digital breast tomosynthesis including synthesized two-dimensional mammograms (DBT) versus digital mammography (DM) in a population-based screening program for breast cancer and to compare selected secondary screening outcomes for the two techniques.

Methods This RCT, performed in Bergen as part of BreastScreen Norway, was approved by the Regional Committees for Medical Health Research Ethics. All screening attendees in Bergen were invited to participate, of which 89% (14,274/15,976) concented during the first year, and were randomized to DBT ($n = 7155$) or DM ($n = 7119$). Secondary screening outcomes were stratified by mammographic density and compared using two-sample t-tests, chi-square tests, ANOVA, negative binomial regression and tests of proportions (z tests).

Results Mean reading time was 1 min 11 s for DBT and 41 s for DM ($p < 0.01$). Mean time spent at consensus was 3 min 12 s for DBT and 2 min 12 s for DM ($p < 0.01$), while the rate of cases discussed at consensus was 6.4% and 7.4%, respectively for DBT and DM ($p = 0.03$). The recall rate was 3.0% for DBT and 3.6% for DM ($p = 0.03$). For women with non-dense breasts, recall rate was 2.2% for DBT versus 3.4% for DM ($p = 0.04$). The rate did not differ for women with dense breasts (3.6% for both). Mean glandular dose per examination was 2.96 mGy for DBT and 2.95 mGy for DM ($p = 0.433$).

Conclusions Interim analysis of a screening RCT showed that DBT took longer to read than DM, but had significantly lower recall rate than DM. We found no differences in radiation dose between the two techniques.

Key Points

- *In this RCT, DBT was associated with longer interpretation time than DM*
- *Recall rates were lower for DBT than for DM*
- *Mean glandular radiation dose did not differ between DBT and DM*

Keywords Mammography · Breast cancer · Mass screening · Digital breast tomosynthesis · Randomized controlled trial

✉ Solveig Hofvind
solveig.hofvind@kreftregisteret.no

[1] Department of Radiology, Haukeland University Hospital, 5021 Bergen, Norway

[2] Department of Clinical Medicine, University of Bergen, 5020 Bergen, Norway

[3] Cancer Registry of Norway, P.O. 5313, 0304 Majorstuen, Oslo, Norway

[4] Sydney School of Public Health, Sydney Medical School, University of Sydney, Camperdown, Australia

[5] Oslo Metropolitan University, Oslo, Norway

Abbreviations

CC	Craniocaudal
DBT	Digital breast thomosynthesis
DM	Digital mammography
MGD	Mean glandular dose
MLO	Mediolateral oblique
RCT	Randomized controlled trial
SD	Standard deviation
SM	Synthetic two-dimensional mammogram
VBD	Volumetric breast density
VDG	Volpara density grade

Introduction

Digital breast tomosynthesis (DBT) in combination with digital mammography (DM) has been claimed to be superior to DM alone in prospective studies of cancer detection in European breast cancer screening programs [1–4]. However, recall rates have been shown to vary between studies.

Globally, a limited number of studies using DBT for screening have reported complete data on interval breast cancers [5–7], and there is presently limited knowledge about the characteristics of the cancers detected with DBT versus DM [5, 7–9]. Further, most studies have evaluated results of DBT in addition to DM, which substantially increases the radiation dose [10–12]. Replacing the DM in DBT + DM with synthetic mammograms (SM), a 2D mammographic image reconstructed from the projection data obtained during the DBT acquisition, has been suggested as a solution and has recently shown promising results with respect to early performance measures in European screening programs [3, 8, 9, 13]. In addition, the sensitivity of DBT among women with dense breasts has been questioned [14–16].

Logistical aspects including increased examination and reading times, the burden on IT systems related to storage, power and speed, and the financial costs are additional aspects that need to be explored to fully evaluate the cost-effectiveness and feasibility of using DBT + SM in organized screening programs.

To address some of the aforementioned gaps in knowledge, we conducted a randomized controlled trial (RCT) using DBT + SM versus DM only: the *To*mosynthesis trial in *Be*rgen (the To-Be trial). The To-Be RCT started in January 2016 and spanned one screening round (2 years). Our study objectives for this paper were to describe the design of this RCT and to report results of interim analyses after the first year of the trial. We compared selected secondary screening outcomes, such as examination time, time spent on screen reading and consensus, rates of cases discussed at consensus, recall rates due to abnormal mammographic findings, and mean glandular dose for DBT + SM (hereafter referred to as DBT) and DM, by mammographic density.

Material and methods

The To-Be trial is approved by the Regional Committees for Medical and Health Research Ethics and registered at ClinicalTrials.gov (NCT02835625).

Study design of RCT

The To-Be trial is an RCT aimed at investigating early performance measures and economical aspects when using DBT versus DM in a screening program for breast cancer (Fig. 1). The trial was performed in Bergen, as a part of BreastScreen Norway, a population-based breast cancer screening program targeting women aged 50–69 years. The program is administered by the Cancer Registry of Norway and has been run since 1995. The program is described in detail elsewhere [17].

All women who attended screening at the screening unit in Bergen, 2016 and 2017, received a request about participation in the trial. Those who agreed and signed an individual consent form were randomized to screening with either DBT or DM, using a 1:1 allocation ratio. The target group for the screening site in Bergen counted about 45,000 women for the actual screening round. Assuming an attendance rate of 75% and 90% participation in the trial, the RCT was powered to identify a statistically significant increase of 25–30% in the rate of screen-detected breast cancers. Information related to the screening examination (screening outcome, procedures performed during recall, mammographic features including density, histologic tumor characteristics, treatment etc.) were reported continuously to the Cancer Registry of Norway by the Breast Center at Haukeland University Hospital in Bergen. Participants will be followed for 2 years after screening, to identify interval breast cancers and cancers in the next screening round.

To avoid bias in the performance of the trial, no results of the surveillance or the analyses, except screening attendance rate and participation rate in the trial, were communicated to the professionals who worked in the practical part of the trial.

Study setting

The To-Be trial was performed in an everyday screening setting. All women underwent standard two-view (craniocaudal and mediolateral oblique views) DBT or DM performed by two radiographers. We used imaging equipment from GE (SenoClaire 3D Breast Tomosynthesis™). The DBT acquisition consisted of nine low-dose exposures over an angle of 25°, reconstructed into 1-mm and 10-mm planes, as well as SM. Screen reading was performed on IDI workstations, each with two 5-megapixel monitors (GE Healthcare MammoWorkstation Version 4.7.0 Image Diagnost). The storage requirement for the raw data and processed image data was 500–3000 MB per examination for DBT and 60–80 MB for DM.

Screening examinations were read using independent double reading. Prior DM screening mammograms were available for subsequently screened women. The standard reading protocol included two views of each breast for DM and two-view synthethic mammograms and 1-mm planes of each breast for DBT. Slabs were vailable for DBT and used in challenging cases, mainly during the consensus meetings. Each breast was assigned a score of 1–5 by each radiologist. A score of 1 indicated screening examination negative for abnormality;

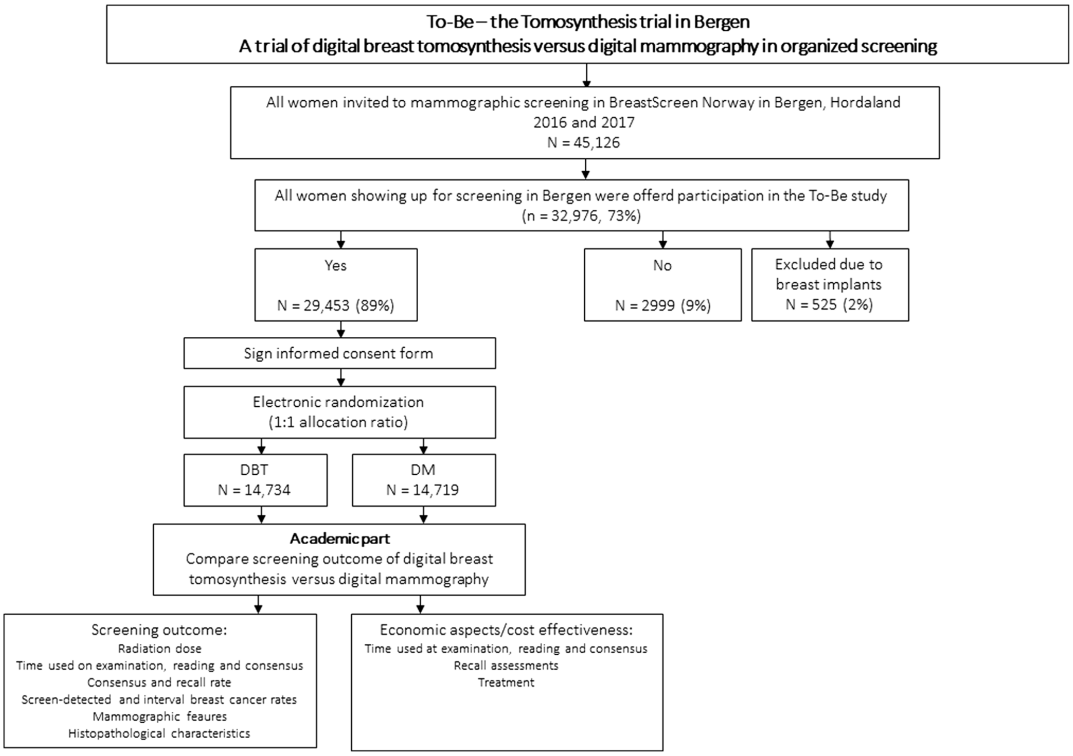

Fig. 1 Study design of the To-Be trial in Bergen, a randomized controlled trial using digital breast tomosynthesis in combination with synthesized 2D images (DBT) versus digital mammography (DM), in Breast Screen Norway. Excluded because of a lack of data on mammographic density

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 to one or both breasts, a consensus meeting (hereafter referred to as consensus) with two or more radiologists was held to determine whether to call the woman back for further assessment (recall).

Up to four prior examinations were available at the workstation both for initial screen reading and consensus. Assessment of recalled women included additional mammographic imaging and/or ultrasound, potentially a needle biopsy and sometimes an MRI. Recall assessment took place at the Breast Center at Haukeland University Hospital.

Eight radiologists with 0–19 years of experience in screen film and/or digital mammography (mean 7 years) took part in screen reading, consensus and follow-up assessments (Appendix, Table 5). All radiologists who did screen reading also performed the assessments for recalled women and diagnostic examinations. DBT was available as a diagnostic method at the Breast Center for about 1 year prior to starting the trial, but had not been used for screening. All radiologists attended a training session with DBT before they started screen reading in the trial. Moreover, a pilot study performed 8 weeks pre-trial included about 300 DBT screening cases.

Study population of interim analyses

These first results from the To-Be trial reports pre-planned interim analyses of selected secondary outcome measures from the first year of To-Be, 2016. A total of 21,786 women were invited to screening in Bergen, whereas 15,976 (73%) attended and 14,274 (89%) agreed to participate in the trial. Altogether, 7155 women were randomized to DBT and 7119 to DM (Fig. 2).

Definition of secondary outcome measures

Examination time was measured as the time spent from when the woman entered the examination room until she left: time was manually registered using a stopwatch for 438 and 535 randomly selected women screened with DBT and DM, respectively, during March 2017.

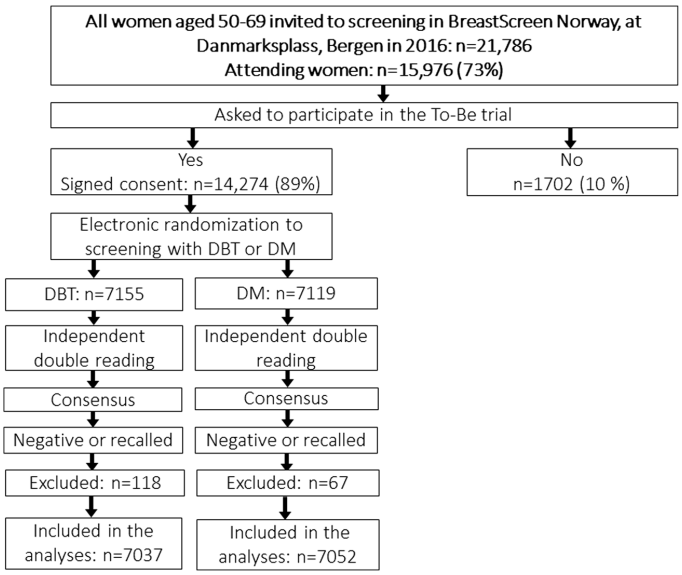

Fig. 2 Study design and study population for interim analyses after 1 year of running the To-Be trial in Bergen, 2016

Time spent on initial screen reading and consensus was measured from the time the radiologist entered the women's ID on the computer until the result of the reading/consensus was registered, using software developed for the trial. Initial screen reading time was measured for each radiologist, while consensus time was measured per woman without taking the number of participating radiologists into account.

The consensus rate was defined as the number of screening examinations discussed at consensus, divided by the total number of screening examinations. For each radiologist, the rate was estimated as the number of examinations given a score of 2 or higher (2+) divided by the number of screen readings per radiologist. The recall rate was the number of women recalled (post-consensus) because of abnormal mammographic findings divided by the number of women screened. For each radiologist, the recall rate was estimated as the number of cases they had read which were discussed at consensus and recalled divided by the number of screen readings by that radiologist.

Measures of volumetric breast density (VBD) and mean radiation dose per exposure (mean glandular dose, MGD) were calculated from the raw image data and data extracted from the DICOM header, using automated software (Volpara version 1.5.1, Volpara Health Technologies Ltd, Wellington, NZ) [18]. Average MGD per screening examination was calculated as the sum of the radiation doses reported by the software for both views and breasts divided by two. VBD was classified into a Volpara density grade (VDG) based on the following scale outlined by Volpara [19]: VDG 1 (VBD < 4.49%); VDG 2 (4.5–7.49%); VDG 3 (VBD 7.5–15.49%) and VDG 4 (VBD ≥ 15.5%). These categories are analogous to the BI-RADS 5th edition density categories a–d [20–22].

Statistical analysis

We estimated mean and median time for screening examination, screen reading and consensus in minutes and seconds (minutes:seconds). For screen reading we excluded outlier values above 10 min and for consensus values above 15 min, assuming that radiologists had been interrupted. The outliers occurred similarly for DBT and DM. Further, we calculated mean values of MGD per examination. Rates of consensus and recall were presented per 100 screening examinations with 95% confidence intervals (95% CI). Analyses were stratified by screening technique (DBT and DM), screening history (prevalent or subsequent attendance), time since trial commencement (1–4, 5–8 and 9–12 months), the radiologists' expertise in screen reading of DM before the start of the trial, and by cumulative number of DM and DBT screen reads in the trial, and mammographic density (VDG 1–4).

Trends in consensus and recall rates according to reading volume were tested for by a negative binomial regression model. We also used negative binomial regression to estimate the risk ratio (RR) and 95% confidence interval (95% CI) of consensus and recall for DBT using DM as the reference. Crude and adjusted RRs were calculated. Covariates in the adjusted models included mammographic density and an interaction effect between screening technique and density.

We used STATA version 15 (Stata Corp, TX) for all statistical analyses and tested differences across categories for

statistical significance using two-sample *t*-tests, chi-square tests, ANOVA and tests of proportions (z test). A *p* value of less than 0.05 was considered statistically significant.

Results

Among women included in the interim analyses, 1% (185/14,274) were excluded because of missing mammographic density data. Information from 14,089 women was thus included in analyses: 7037 screened with DBT and 7052 screened with DM. Women were, on average, 59 years old at screening in both groups ($p = 0.469$) (Table 1). The distribution of characteristics detailed in Table 1 did not differ between the two groups.

Women spent an average time (minutes:seconds) of 5:24 (median 5:13) for DBT and 4:19 (median 4:07) for DM in the screening examination room ($p < 0.01$) (Table 2). Average and median times spent on initial screen reading and consensus were generally higher for DBT compared to DM.

The rates of cases discussed at consensus were 6.4% for DBT and 7.4% for DM ($p = 0.03$) (Table 3). These rates did not differ among prevalent examinations (13.0% for both DBT and DM, $p = 0.97$), which was in contrast to the subsequent examinations, where the rate was 5.2% for DBT and 6.3% for DM ($p < 0.01$). We observed an increasing rate of cases discussed at consensus by VDG for DBT (p for trend < 0.01), but not for DM (p for trend = 0.078).

The eight radiologists' reading volume before and during the trial period varied (Appendix, Table 5). A score of 2+, resulting in a consensus meeting, was given for an average of 4.5% of the DBT and 5.4% of the DM screen reads for each of the radiologists (Appendix, Table 6). The consensus rate decreased with 0.1% for DBT ($p = 0.4$) and 0.2% for DM ($p = 0.05$) per 1000 DM screen reads prior to start-up of the trial.

The recall rate was 3.0% for DBT and 3.6% for DM ($p = 0.03$) (Table 3). This rate did not differ for the two techniques among prevalently screened women (6.3% for DBT and 6.2% for DM, $p = 0.95$), in contrast to subsequently screened women where the rate was 2.3% for DBT and 3.1% for DM ($p < 0.01$). For DBT, recall rates increased from 2.2% for women with VDG 1 to 3.6% for women with VDG 4 (p for trend < 0.01). No statistically significant difference was observed for women screened with DM ($p = 0.93$). The number of DM screen reads before the trial period did not significantly alter the recall rates for DBT or DM ($p = 0.6$ for DBT and $p = 0.8$ for DM) (Appendix Table 5).

The cumulative reading volume of DBT during the trial showed a non-significant trend of a decreasing consensus rate (RR = 0.95, $p = 0.3$) (Appendix Table 6 and Fig. 4). For DM, this trend reached statistical significance (RR = 0.93, $p = 0.04$). For recall rates, a non-significant trend of decreasing value with cumulative reading volume during the trial period was observed both for DBT and DM ($p = 0.8$ for DBT and $p = 0.4$ for DM).

Table 1 Characteristics of the study population screened with digital breast tomosynthesis including synthesized 2D mammography (DBT) or digital mammography (DM) in the To-Be trial in Bergen, 2016

	DBT ($n = 7037$)	DM ($n = 7052$)	*p* value
Age (years)			
Mean/median	59/59	59/59	0.469*
50–54	27.6%	27.6%	0.983**
55–59	25.5%	25.8%	
60–64	24.9%	24.7%	
65–71	22.0%	21.9%	
Screening history (% of screened women)			0.883**
Prevalently screened	15.7%	15.6%	
Subsequently screened	84.4%	84.4%	
Mammographic density			0.248**
VDG 1	21.0%	20.4%	
VDG 2	44.8%	43.7%	
VDG 3	26.1%	27.1%	
VDG 4	8.2%	8.8%	

*t-test for means

**Chi-square test

The adjusted risks of consensus and recall were lower for DBT than for DM: RR 0.71 (95% CI 0.52–0.97) for consensus and 0.58 (95% CI 0.38–0.89) for recalls (Table 4). The interaction between screening technique and mammographic density was not stastistically significant when modelling the risk of consensus. However, the risk of recall among women screened with DBT increased for VDG 3 versus VDG 1 ($p = 0.033$), and displayed a trend toward increased values for VDG 4 versus VDG 1 ($p = 0.061$), compared with DM.

MGD per examination was 2.96 mGy for DBT and 2.95 mGy for DM ($p = 0.433$) (Fig. 3). It did not differ with mammographic density, nor within the density groups or between screening techniques.

Discussion

In the first year of this RCT using DBT and DM in population-based breast cancer screening, we found lower consensus and recall rates among women screened with DBT than with DM. Our density-stratified analyses identified that recall rates were lower for DBT only for women with non-dense breasts (VDG 1 and VDG 2). Time spent both on screen reading and consensus was longer for DBT than for DM. Average MGD did not differ between the two techniques.

The lower recall rate for DBT compared to DM found in our interim analyses supports results from other studies,

Table 2 Mean and median time spent in the examination room per woman, at initial screen reading per radiologist, and at consensus for digital breast tomosynthesis with synthesized 2D (DBT) versus digital mammography (DM), in the To-Be trial in Bergen, 2016

	DBT	DM	*p* value*
Examination time per woman	*N* = 438	*N* = 534	
Mean/median (min:s)	5:24/5:13	4:19/4:07	< 0.01
Initial screen reading time per reader (min:s)	*N* = 7029	*N* = 7048	
All screens	1:11/0:54	0:41/0:26	< 0.01
Prevalent screens	1:10/0:53	0:33/0:19	< 0.01
Subsequent screens	1:11/0:54	0:43/0.27	< 0.01
p for trend*	0.850	< 0.01	
Reading time stratified by time since start of trial			
1–4 months	1:18/1:00	0:42/0:29	< 0.01
5–8 months	0:56/0.46	0:33/0:21	< 0.01
9–12 months	1:11/0.54	0:45/0:27	< 0.01
p for trend**	< 0.001	< 0.001	
Reading time stratified by mammographic density			
VDG 1	1:01/0:47	0:39/0:24	< 0.01
VDG 2	1:09/0:55	0:40/0:26	< 0.01
VDG 3	1:15/0:58	0:44/0:28	< 0.01
VDG 4	1:17/0:58	0:42/0:28	< 0.01
p for trend**	< 0.001	< 0.001	
Time spent on consensus (min:s)	*N* = 451	*N* = 519	
All	3:12/2:42	2:12/1:55	< 0.01
Prevalent screens	2:51/2:27	1:51/1:36	< 0.01
Subsequent screens	3:22/2:49	2:20/2:04	< 0.01
p for trend*	< 0.001	< 0.001	
Consensus time stratified by time since start of trial			
1–4 months	3:31/3:14	2:08/1:48	< 0.01
5–8 months	2:45/2:14	1:54/1:42	< 0.01
9–12 months	3:06/2:39	2:21/2:05	< 0.01
p for trend**	0.012	0.014	
Consensus time stratified by mammographic density			
VDG 1	3:15/2:33	2:15/2:03	< 0.01
VDG 2	3:14/2:47	2:12/1:51	< 0.01
VDG 3	3:16/2:48	2:14/1:56	< 0.01
VDG 4	2:52/2:30	2:00/1:51	< 0.01
p for trend**	0.623	0.695	

*t-test for means

**ANOVA

although recall rates have been shown to vary [1–4, 8, 9]. Different reading protocols and screening logistics might be some of the reasons for this variance [23–26]. Reducing recall rates below 3% in organized screening programs seems more challenging than reducing a recall rate of 10% or higher. Regardless of screening technique, there is limited evidence on what the ideal recall rate is, according to false positive screening results, cancer detection and breast cancer mortality [27, 28].

More than 65% of the women in our study were classified as having non-dense breast (VDG 1 or VDG 2). Women with non-dense breasts had a lower recall rate when screened with DBT than when screened with DM. However, recall rates did not differ between DBT and DM for women with dense breasts (VDG 3 or VDG4). Moreover, the effect of mammographic density on the risk of recall tended to be larger for DBT than for DM, a relevant finding in a breast cancer screening program given that it applies to the larger proportion of screening attendees in our population. Given the established knowledge about the increasing risk of breast cancer with mammographic density, the increase in recall rate with density seems reasonable.

The consenus rates were also higher for women with dense rather than fatty breasts, both for DBT and DM. This is possibly related to the complex parenchyma and the need for a

Table 3 Numbers (*n*) and percentages (%) of screening examinations discussed at consensus and recalls for digital breast tomosynthesis with synthesized 2D (DBT) versus digital mammography (DM), in the To-Be trial in Bergen, 2016

	Discussed at consensus			Recalled		
	DBT (*n* = 7037) *N* % (95% CI)	DM (*n* = 7052) *N* % (95% CI)	*p* value**	DBT (*n* = 7037) *N* % (95% CI)	DM (*n* = 7052) *N* % (95% CI)	*p* value**
All screens	451/7037 6.4% (5.8–7.0)	519/7052 7.4% (6.8–8.0)	0.03	208/7037 3.0% (2.6–3.4)	254/7052 3.6% (3.2–4.0)	0.03
Prevalent screens	143/1101 13.0% (11.0–15.0)	143/1097 13.0 (11.0–15.0)	0.97	69/1101 6.3% (4.8–7.7)	68/1097 6.2% (4.8–7.6)	0.95
Subsequent screens	308/5936 5.2% (4.6–5.8)	376/5955 6.3% (5.7–6.9)	< 0.01	139/5936 2.3% (2.0–2.7)	186/5955 3.1% (2.7–3.6)	< 0.01
p for trend*	< 0.01	< 0.01		< 0.01	< 0.01	
Time since start of trial						
1–4 months	175/2676 6.5% (5.6–7.5)	190/2641 7.2% (6.2–8.2)	0.35	81/2676 3.0% (2.4–3.7)	95/2641 3.6% (2.9–4.3)	0.25
5–8 months	76/1431 5.3% (4.2–6.5)	83/1463 5.7% (4.5–6.9)	0.67	37/1431 2.6% (1.8–3.4)	29/1463 2.0% (1.3–2.7)	0.28
9–12 months	200/2930 6.8% (5.9–7.7)	246/2948 8.3% (7.3–9.3)	0.03	90/2930 3.1% (2.4–3.7)	130/2948 4.4% (3.7–5.2)	< 0.01
p for trend*	0.149	< 0.01		0.648	< 0.01	
Mammographic density						
VDG 1	63/1475 4.3% (3.2–5.3)	87/1441 6.0% (4.8–7.3)	0.03	32/1475 2.2% (1.4–2.9)	49/1441 3.4% (2.5–4.3)	0.04
VDG 2	189/3150 6.0% (5.2–6.8)	224/3082 7.3% (6.4–8.2)	0.04	78/3150 2.5% (1.9–3.0)	110/3082 3.6% (2.9–4.2)	0.01
VDG 3	148/1836 8.1% (6.8–9.3)	154/1910 8.1% (6.8–9.3)	1.0	77/1836 4.2% (3.3–5.1)	73/1910 3.8% (3.0–4.7)	0.56
VDG 4	51/576 8.9% (6.5–11.2)	54/619 8.7% (6.5–10.9)	0.94	21/576 3.6% (2.1–5.2)	22/619 3.6% (2.1–5.0)	0.93
p for trend*	< 0.01	0.078		< 0.01	0.93	

**t*-test for means

***ANOVA*

second opinion. The consensus meeting used in BreastScreen Norway can be considered an educational activity where "positive" cases are discussed and prior screening exams are carefully considered before a final decision about recall is made. In a broader perspective, our results, demonstrating a lower percentage of cases needing to be discussed at consensus, suggest that DBT may reduce the percentage of cases needing third arbitrating reads in other programs. As far as we know, no other studies have reported consensus rates for DBT previously. It is possible that the dense cases discussed at consensus were more obvious to recall than the fatty cases. The radiologists might thus need less time to agree about recall for the dense versus the fatty cases.

The burden of increased examination and screen reading time from DBT is a critical issue for screening programs. The increased examination time was mainly due to time spent on explaining to the women how the x-ray machine would move and to make the x-ray tube ready for exposure. This extra time is expected to be reduced or resolved in subsequent screening rounds. We demonstrated that the average reading time was 30 s longer for DBT than for DM at initial screen reading (1:11 versus 0:41, respectively). The Oslo Tomosynthesis Screening Trial (OTST) reported that an additional 41 s was needed for reading DBT compared to DM [2], while results from the STORM trial, Malmo trial and a study by Dang et al showed an increase of 44 s [1], 30 s [4] and 54 s [29], respectively. Our results therefore represent the minimum increase in time spent on initial screen reading reported in the literature to date. However, the reading time varied between radiologists. We found that some radiologists were fast readers while other used more time. We consider this variability amongst the radiologists as individual-related rather than trial-related since the findings were independent of screening technique and volume of screen reads during their career.

In our study, time spent on screen reading and consensus was lowest 5–8 months after the start of the trial. This could be because this period was during the summer months, when fewer women were screened, resulting in low power in the estimate. The low reading and consensus time could also be related to a learning effect. A workshop reviewing

Table 4 Risk ratio (RR) of undergoing consensus and being recalled adjusted for mammographic density for digital breast tomosynthesis with synthesized 2D (DBT) versus digital mammography (DM) in the To-Be trial in Bergen, 2016

	RR of consensus			RR of recall		
	RR	95% CI	*p* value	RR	95% CI	*p* value
Screening technique						
DM	1.00	–	–	1.00	–	–
DBT	0.71	(0.52–0.97)	0.032	0.58	(0.38–0.89)	0.013
Mammographic density						
VDG 1	1.00	–	–	1.00	–	–
VDG 2	1.20	(0.95–1.53)	0.129	1.00	(0.73–1.37)	0.979
VDG 3	1.34	(1.04–1.72)	0.025	1.14	(0.81–1.59)	0.472
VDG 4	1.44	(1.04–2.00)	0.027	1.08	(0.68–1.72)	0.752
Screening technique and mammographic density (interaction)						
DBT–VDG 1	1.00	–	–	1.00	–	–
DBT–VDG 2	1.17	(0.81–1.68)	0.410	1.31	(0.79–2.18)	0.302
DBT–VDG 3	1.41	(0.96–2.07)	0.077	1.77	(1.05–3.01)	0.033
DBT–VDG 4	1.43	(0.88–2.33)	0.143	1.93	(0.97–3.84)	0.061

cancer cases dismissed by one of the two readers was performed 7–8 months after the start of the trial, as a part of the usual quality assurance in the program. This might have contributed to readers deliberating longer at screen reading and may account for the increased reading time in the third period, 8–12 months after trial commencement.

Results from other studies indicate the need for training and workshops before reading DBT in screening [30, 31]. In our study, the radiologists' experience in DM screen reading before the trial period varied from beginners to very experienced, the latter with more than 100,000 screen reads during their career as a breast radiologist. Not all radiologists participated in screen reading DBT in the pilot, which was performed 8 weeks before the trial commenced. We identified a significant decreasing trend of consensus with reading volume during the trial for DM, but not for DBT. The volume of screen reads prior to the trial did not show any correlation with either consensus or recall rate, neither for DBT nor DM. Our study presents results only for the first year of the trial, which might be considered the learning period. Further analyses including a longer study period might shed a different light on the issue. In this trial radiologists without experience in screen reading did training on test sets, shadow reading within the trial and performed clinical mammography with DBT. In retrospect, the pilot could have been extended to 6 months to enhance reader preparation, and additional workshops could have been held to make sure all participating radiologists had read a minimum number of negative and false positive examinations, screen-detected and

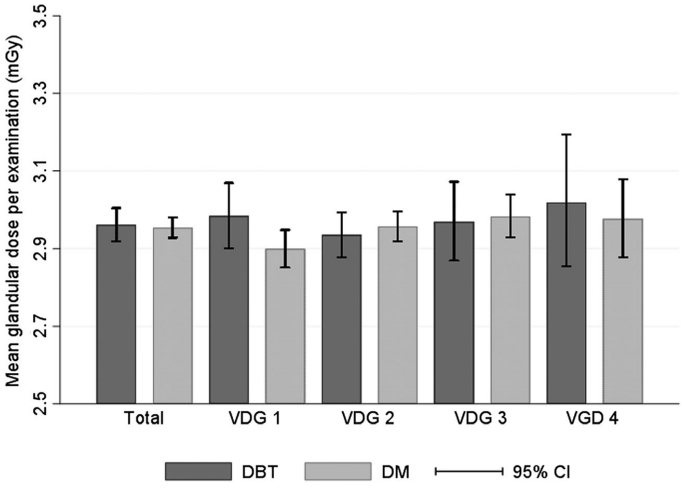

Fig. 3 Mean glandular dose (MGD) per examination among women screened during the first year of the To-Be trial, overall and by Volpara density grade (VDG), stratified by imaging technique (digital breast tomosynthesis including synthesized 2D mammograms [DBT] or digital mammography [DM])

interval breast cancers before the trial started. Although a roster was established at the start of the trial to ensure all radiologists read equal numbers of DBT and DM cases, this plan was not strictly followed because of varying individual work speeds and an unforeseen high volume of mammography outside of the screening program. Moreover, participating radiologists were not all exposed to the same number of DBT cases. The issues encountered in the implementation of the To-Be trial represent real-world screening challenges and provide novel insights that should inform other breast screening programs when planning DBT evaluations.

We found no statistically significant difference in radiation dose per examination between DBT and DM. Gennaro et al [10] reported doses per view (CC, MLO), also calculated by Volpara, for examinations acquired using a different unit/system and found the doses to be statistically significantly higher for DBT than for DM for both views. In a per view comparison (DBT and DM exposures of the same breasts during the same compression session) they found an average increase in DBT dose compared to DM of 38% (range 0–75%). Similarly, the Oslo Tomosynthesis Screening Trial used DBT systems from the same vendor as Gennaro et al and found, on average, dose per view to be 23% higher with DBT than DM when machine-reported doses were compared [12].

Using a system from yet another manufacturer, Lång et al [4] did not report dose values; instead, the automatic exposure control was set to yield an average dose of 1.2 mGy for DM and 1.6 mGy for DBT for a standard breast model. This gives an expected per view ratio of MGD_{DBT}/MGD_{DM} of 1.33. For our system the manufacturer stated that the target MGD for DBT using automatic exposure control was equivalent to the MGD per view for DM, i.e. an expected ratio of approximately 1. The absence of a difference between MGD with DBT and DM observed in our study is therefore in line with how the system is set to operate by the manufacturer.

During the study period, routine quality assurance of the collected data and control activities were performed. We consider this to be one important strength of this study. We used an RCT design, the most reliable research design to compare screening modalities, and embedded this in a population-based screening program; these features of our trial minimize bias and increase the generalizability to other organized screening programs.

A limitation of this study is the short time spent on training and workshops in DBT for radiographers before the start of the trial, which could have influenced the results in either direction [30, 31]. Moreover, we have not presented breast cancer detection data; this decision was based on per protocol power estimation, which showed that 2 years of screening—one screening round—was needed to show a 25–30% difference in the rate of screen-detected breast cancer between DBT and DM. The moderate number of cases included in the analyses also represents a limitation in this study, particularly when stratifying into subgroups. Despite these limitations, we present our interim results to inform other population-based screening programs of selected secondary screening outcomes from an RCT of DBT and DM, in particular the estimated recall rate, screen reading time and radiation metrics, all of which matter to screening practice and research planning. To the best of our knowledge, there are no published secondary screening outcomes from other RCTs of DBT screening.

In conclusion, after the first year of running an RCT comparing DBT and DM, including about 7000 screened women in each arm, we showed a lower recall rate for women screened with DBT than DM. Our RCT sheds further light on the burdens of interpretation time and radiation dose, which are key factors in population-based screening. Time spent on screen reading and on consensus was longer for DBT than for DM. MGD measured by automated software on a GE SenoClair machine did not differ between the two techniques. Our results are somewhat different from other published studies and call for RCTs from different screening populations and with equipment from different vendors in order to gain evidence about the consequences of implementing DBT with synthesized mammograms, as a screening technique in population-based screening programs.

Acknowledgements We would like to thank all the radiographers, radiologists and others at the Breast Center, Department of Radiology at Haukeland University Hospital, Bergen, Norway, for their extensive work, help and support in the performance of this study. Solveig Hofvind is the principal investigator for the study and head of BreastScreen Norway. She is employed at the Cancer Registry of Norway, which administers the screening program. Nehmat Houssami is funded by a Breast Cancer Research Leadership Fellowship from the National Breast Cancer Foundation (Australia).

Funding This study has received funding by the Research Council of Norway (Project # 247941).

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Solveig Hofvind.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry No complex statistical methods were necessary for this paper.

Informed consent Written informed consent was obtained from all subjects (patients) in this study.

Ethical approval Institutional review board approval was obtained.

Methodology
- prospective
- randomized controlled trial
- performed at one institution

Appendix

Table 5 Characteristics of the radiologists involved in the To-Be trial in Bergen, screen-reads (n), rates of consensus (score 2+) and recalls for DBT and DM by radiologist

Radiologist	Age	Started screen-reading in BreastScreen Norway (month, year)	Started screen-reading DBT in the trial (month, year)	DM screen-readings before the trial period (n)	Screen reads (n)		Score 2+ Consensus (%)		Recall (%)	
					DBT	DM	DBT	DM	DBT	DM
R1	36	Feb 2016	Feb 2016	0	2978	3884	4.1 %	5.3 %	2.4 %	3.1 %
R2	32	May 2014	Apr 2016	10744	920	383	7.1 %	8.1 %	3.8 %	3.7 %
R3	47	Oct 2010	Jan 2016	15085	1781	1344	4.3 %	4.9 %	3.1 %	3.8 %
R4	36	May 2012	Jan 2016	23801	1208	1563	4.4 %	6.7 %	3.1 %	4.7 %
R5	50	Jan 2009	Aug 2016	24015	453	502	4.6 %	9.2 %	3.1 %	4.4 %
R6	43	Sept 2007	Jan 2016	37361	1634	1177	4.8 %	5.2 %	3.1 %	3.8 %
R7	40	Sept 2008	Jan 2016	92590	2155	3789	4.2 %	4.0 %	2.7 %	3.0 %
R8	50	Sept 1997	Jan 2016	109152	2945	1462	4.4 %	6.6 %	3.2 %	4.9 %
p for trend							*0.4*	*0.05*	*0.6*	*0.8*
Total					14 074	14 104	4.5 %	5.4 %	3.0 %	3.6 %

p for trend tested by a negative binomial regression model

Table 6 Cumulative number of DBT and DM screen-reads during the first year of the To-Be trial in Bergen and subsequent rates of consensus and recalls

Number of DBT screen-reads in To-Be before current reading	Screening examinations (n)	Consensus (%)	Recall (%)
0-499	3955	4.8 %	3.2 %
500-999	3420	4.8 %	2.8 %
1000-1499	2707	4.5 %	2.9 %
1500-1999	1914	3.4 %	2.7 %
2000-2499	1155	4.7 %	3.1 %
2500+	923	4.4 %	3.0 %
p for trend		*0.3*	*0.8*
Total DBT	14074	4.5 %	3.0 %
Number of DM screen-reads in To-Be before current reading			
0-499	3922	6,0 %	4,1 %
500-999	3002	5,6 %	3,6 %
1000-1499	2446	5,9 %	4,2 %
1500-1999	1063	3,4 %	1,5 %
2000-2499	1000	4,3 %	2,2 %
2500+	2671	5,0 %	3,8 %
p for trend		*0.04*	*0.4*
Total DM	14104	5,4 %	3,6 %

p for trend tested by a negative binomial regression model

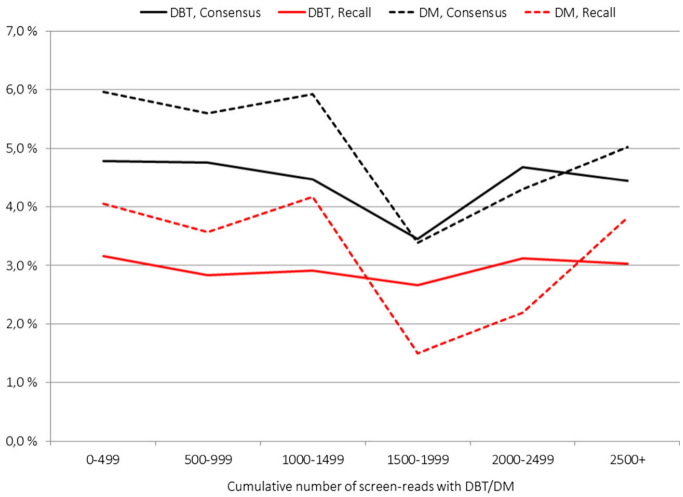

Fig. 4 Rates of consensus and recalls by cumulative number of DM (dotted line) and DBT (solid line) screen-readings during the first year of the To-Be trial in Bergen, 2016

References

1. Ciatto S, Houssami N, Bernardi D et al (2013) Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 14:583–589
2. Skaane P, Bandos AI, Gullien R et al (2013) Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 267:47–56
3. Bernardi D, Macaskill P, Pellegrini M et al (2016) 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. Lancet Oncol 17:1105–1113
4. Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S (2016) 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. Eur Radiol 26:184–190
5. Skaane P, Sebuødegård S, Bandos AI et al (2018) Performance of breast cancer screening using digital breast tomosynthesis: Results from the prospective population-based Oslo Tomosynthesis Screening Trial. Breast Cancer Res Treat 169:489–496
6. Houssami N, Bernardi D, Caumo F et al (2018) Interval breast cancers in the 'screening with tomosynthesis or standard mammography' (STORM) population-based trial. Breast 38:150–153
7. Bahl M, Gaffney S, McCarthy AM, Lowry KP, Dang PA, Lehman CD (2018) Breast cancer characteristics associated with 2D digital mammography versus digital breast tomosynthesis for screening-detected and interval cancers. Radiology 287:49–57
8. Hofvind S, Hovda T, Holen ÅS et al (2018) Digital breast tomosynthesis and synthetic 2D mammography versus digital mammography: evaluation in a population-based screening program. Radiology 287:787–794
9. Caumo F, Zorzi M, Brunelli S et al (2018) Digital breast tomosynthesis with synthesized two-dimensional images versus full-field digital mammography for population screening: outcomes from the Verona screening program. Radiology 287:37–46
10. Gennaro G, Bernardi D, Houssami N (2018) Radiation dose with digital breast tomosynthesis compared to digital mammography: per-view analysis. Eur Radiol 28:573–581
11. Svahn TM, Houssami N, Sechopoulos I, Mattsson S (2015) Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. Breast 24:93–99
12. Østerås BH, Skaane P, Gullien R, Martinsen ACT (2018) Average glandular dose in paired digital mammography and digital breast tomosynthesis acquisitions in a population based screening program: effects of measuring breast density, air kerma and beam quality. Phys Med Biol 63:035006
13. Skaane P, Bandos AI, Eben EB et al (2014) Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. Radiology 271:655–663
14. Rafferty EA, Durand MA, Conant EF et al (2013) Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts. JAMA 15:1784–1786
15. Friedewald SM, Rafferty EA, Rose SL et al (2014) Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 311:2499–2507
16. Gilbert FJ, Tucker L, Young KC (2016) Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol 71:141–150
17. Hofvind S, Tsuruda K, Mangerud G 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; 2017. https://www.kreftregisteret.no/globalassets/cancer-in-

Norway/2016/mammo_cin2016_special_issue_web.pdf Cited 03012018
18. Aitken Z, McCormack VA, Highnam RP et al (2010) Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. Cancer Epidemiol Biomarkers Prev 19: 418–428
19. Highnam R, Brady M, Yaffe M, Karssemeijer N, Harvey J (2010) Robust breast composition measurement - Volpara™. In: Martí JOA, Freixenet J, Martí R (eds) Lecture notes in computer science: 10th international workshop on digital mammography. Springer-Verlag, Berlin Heidelberg, pp 342–349
20. Singh T, Sharma M, Singla V, Khandelwal N (2016) Breast density estimation with fully automated volumetric method: comparison to radiologists' assessment by BI-RADS categories. Acad Radiol 23: 78–83
21. Sartor H, Lång K, Rosso A, Borgquist S, Zackrisson S, Timberg P (2016) Measuring mammographic density: comparing a fully automated volumetric assessment versus European radiologists' qualitative classification. Eur Radiol 26:4354–4360
22. van der Waal D, den Heeten GJ, Pijnappel RM et al (2015) Comparing visually assessed BI-RADS breast density and automated volumetric breast density software: a cross-sectional study in a breast cancer screening setting. PLoS One 10:e0136667
23. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS (2017) Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. Radiology 283:70–76
24. Freer PE, Riegert J, Eisenmenger L et al (2017) Clinical implementation of synthesized mammography with digital breast tomosynthesis in a routine clinical practice. Breast Cancer Res Treat 166:501–509
25. Baltzer PAT, Kapetas P, Marino MA, Clauser P (2017) New diagnostic tools for breast cancer. Memo 10:175–180
26. Houssami N, Hunter K, Zackrisson S (2017) Overview of tomosynthesis (3D mammography) for breast cancer screening. Breast Cancer Manag 6:179–186
27. Yankaskas BC, Schell MJ, Miglioretti DL (2004) Recall and detection rates in screening mammography. Cancer 101:2710–2711 author reply 1-2
28. Schell MJ, Yankaskas BC, Ballard-Barbash R et al (2007) Evidence-based target recall rates for screening mammography. Radiology 243:681–689
29. Dang PA, Freer PE, Humphrey KL, Halpern EF, Rafferty EA (2014) Addition of tomosynthesis to conventional digital mammography: effect on image interpretation time of screening examinations. Radiology 270:49–56
30. Tucker L, Gilbert FJ, Astley SM et al (2017) Does reader performance with digital breast tomosynthesis vary according to experience with two-dimensional mammography? Radiology 283:371–380
31. Zhang J, Grimm LJ, Lo JY et al (2015) Does breast imaging experience during residency translate into improved initial performance in digital breast tomosynthesis? J Am Coll Radiol 12:728–732

Paper 2

ORIGINAL RESEARCH • BREAST IMAGING

Radiology

Comparing Screening Outcomes for Digital Breast Tomosynthesis and Digital Mammography by Automated Breast Density in a Randomized Controlled Trial: Results from the To-Be Trial

Nataliia Moshina, MD, PhD • Hildegunn S. Aase, MD • Anders S. Danielsen, MSc • Ingfrid S. Haldorsen, MD, PhD • Christoph I. Lee, MD • Sophia Zackrisson, MD, PhD • Solveig Hofvind, PhD

From the Cancer Registry of Norway, PO Box 5313, Majorstuen, 0304 Oslo, Norway (N.M., A.S.D., S.H.); Department of Radiology, Haukeland University Hospital, Bergen, Norway (H.S.A., I.S.H.); Department of Clinical Medicine, University of Bergen, Bergen, Norway (H.S.A., I.S.H.); Department of Radiology, Seattle Cancer Care Alliance, University of Washington, Seattle, Wash (C.I.L.); Department of Translational Medicine, Diagnostic Radiology, Lund University Cancer Center, Malmö, Sweden (S.Z.); and Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway (S.H.). Received March 26, 2020; revision requested May 13; revision received July 20; accepted July 30. **Address correspondence to** S.H. (e-mail: *solveig.hofvind@kreftregisteret.no*).

The To-Be trial is jointly funded by the Department of Radiology at Haukeland University Hospital, Cancer Registry of Norway, and the Research Council of Norway.

Conflicts of interest are listed at the end of this article.

See also the editorial by Sechopoulos and Athanasiou in this issue.

Radiology 2020; 297:522–531 • https://doi.org/10.1148/radiol.2020201150 • Content code: BR

Background: Digital breast tomosynthesis (DBT) is considered superior to digital mammography (DM) for women with dense breasts.

Purpose: To identify differences in screening outcomes, including rates of recall, false-positive (FP) findings, biopsy, cancer detection rate, positive predictive value of recalls and biopsies, and histopathologic tumor characteristics by density using DBT combined with two-dimensional synthetic mammography (SM) (hereafter, DBT+SM) versus DM.

Materials and Methods: This randomized controlled trial comparing DBT+SM and DM was performed in Bergen as part of BreastScreen Norway, 2016–2017. Automated software measured density (Volpara Density Grade [VDG], 1–4). The outcomes were compared for DBT+SM versus DM by VDG in descriptive analyses. A stratified log-binomial regression model was used to estimate relative risk of outcomes in subgroups by screening technique.

Results: Data included 28 749 women, 14 380 of whom were screened with DBT+SM and 14 369 of whom were screened with DM (both groups: median age, 59 years; interquartile range [IQR], 54–64 years). The recall rate was lower for women screened with DBT+SM versus those screened with DM for VDG 1 (2.1% [81 of 3929] vs 3.3% [106 of 3212]; $P = .001$) and VDG 2 (3.2% [200 of 6216] vs 4.3% [267 of 6280]; $P = .002$). For DBT+SM, adjusted relative risk of recall (VDG 2: 1.8; $P < .001$; VDG 3: 2.4; $P < .001$; VDG 4: 1.8; $P = .02$) and screen-detected breast cancer (VDG 2: 2.4; $P = .004$; VDG 3: 2.8; $P = .01$; VDG 4: 2.8; $P = .05$) increased with VDG, whereas no differences were observed for DM (relative risk of recall for VDG 2: 1.3; $P = .06$; VDG 3: 1.1; $P = .41$; VDG 4: 1.1; $P = .71$; and relative risk of screen-detected breast cancer for VDG 2: 1.7; $P = .13$; VDG 3: 2.1; $P = .06$; VDG 4: 2.2; $P = .15$).

Conclusion: Screening with digital breast tomosynthesis combined with synthetic two-dimensional mammograms (DBT+SM) versus digital mammography (DM) yielded lower recall rates for women with Volpara Density Grade (VDG) 1 and VDG 2. Adjusted relative risk of recall and screen-detected breast cancer increased with denser breasts for DBT+SM but not for DM.

Online supplemental material is available for this article.

An earlier incorrect version appeared online and in print. This article was corrected on January 27, 2022.

Digital breast tomosynthesis (DBT) increases the incidence of screen-detected breast cancer (SDC) when compared with standard digital mammography (DM) in paired and nonpaired prospective trials and retrospective studies (1–4). Results of a randomized controlled trial performed in Italy using DBT in combination with standard DM versus DM alone also support these findings (1). However, a randomized controlled trial performed by our group in Bergen, as part of BreastScreen Norway, had a different conclusion (5). In the latter trial, we found that DBT that includes synthetic two-dimensional mammography (SM) (hereafter, DBT+SM) yielded a breast cancer detection rate similar to that of DM.

Mammographic density is an independent risk factor for breast cancer (6,7) and is known to mask breast malignancies (8). Mammographic density has been subjectively classified according to Breast Imaging Reporting and Data System assessment for decades (9), despite limitations related to inter- and intrareader agreement (10–12). Automated estimation of mammographic density eliminates subjectivity while increasing reliability. Thus, it is the preferred method for measuring density in European breast cancer screening programs (13,14). We have previously documented a sensitivity of 70% for women in the highest versus lowest automated density category (70% vs 86%, respectively) in BreastScreen Norway

Abbreviations

BMI = body mass index, CI = confidence interval, DBT = digital breast tomosynthesis, DM = digital mammography, FP = false positive, IQR = interquartile range, RR = relative risk, SDC = screen-detected breast cancer, SM = synthetic mammogramphy, VBD = volumetric breast density, VDG = Volpara Density Grade

Summary

The relative risks of recall and screen-detected breast cancer increased by automated breast density category 1–4 for digital breast tomosynthesis combined with synthetic mammograms but not for digital mammography.

Key Results

- Using automated breast density software, women with nondense breasts had a lower recall rate when screened with digital breast tomosynthesis combined with synthetic mammography (DBT+SM) than with standard digital mammography (DM) alone (Volpara Density Grade [VDG] 1: 2.1% vs 3.3%; $P = .001$; VDG 2: 3.2% vs 4.3%; $P = .002$).
- Regardless of breast density, the rate of screen-detected breast cancer did not differ between DBT+SM and DM (VDG 1: 0.46% vs 0.47%; $P = .96$; VDG 2: 0.77% vs 0.62%; $P = .31$; VDG 3: 0.73% vs 0.68%; $P = .82$; VDG 4: 0.62% vs 0.61%; $P = .98$).

when dividing women into four categories of volumetric breast density (15).

Currently, there are few reports regarding DBT+SM versus DM screening performance by volumetric breast density in a population-based screening program (16). As part of the To-Be trial, which randomly assigned women to DBT+SM or DM screening, we collected information on volumetric breast density using automated software. The objective of this stratified analysis was to identify differences in recall, false-positive (FP) screening examinations, and biopsy rates; SDC; and histopathologic tumor characteristics for DBT+SM versus DM screening by automated measured mammographic density. We hypothesized that DBT+SM would have superior screening performance compared with DM in women with high automated volumetric density.

Materials and Methods

The prospective randomized controlled trial (NCT02835625), including this secondary analysis, was approved by the Regional Committee for Medical and Health Research Ethics in the South East of Norway (2015/424). The study did not receive any support from industry. Data generated or analyzed during the study are available from the corresponding author, by request.

Study Design and Participants

The trial was embedded within the population-based breast cancer screening program BreastScreen Norway, 2016–2017 (17). The DBT acquisition consisted of nine exposures reconstructed into SM (5). Independent double reading with consensus, according to usual procedures in the program, was performed. Further details on BreastScreen Norway and the To-Be trial are described elsewhere (5,17,18).

Women participating in the To-Be trial were assigned to DBT+SM or DM by using simple random allocation after providing a signed written consent form. In women diagnosed with more than one breast cancer, we used a hierarchy of severity to define which cancer was to be included in the analyses. All examinations were performed with GE Senographe Essential SenoClaire (GE Healthcare, Chicago, Ill). Image Diagnost International Workstations from GE Healthcare were used for interpretation by eight radiologists (including H.S.A.) with varying levels of experience in screen-reading DBT+SM and DM (5).

Two studies published data from the trial: one interim analysis that included 7089 women as well as an article that analyzed the primary outcome and included 28749 women (5,18). Information on density was included in the interim analysis (18).

Data and Mammographic Density Measurements

Women were categorized by breast density. An automated software (VolparaDensity, version 1.5.4; *http://www.volparasolutions.com/our-products/volparadensity/*) (19) was integrated in the picture archiving and communication system. A density grade (ie, Volpara Density Grade [VDG], 1–4) that is analogous to the four-category Breast Imaging Reporting and Data System (5th edition) classification was obtained from the DM image or the central projection of the DBT slices (9). Volumetric assessment in the study differs from the subjective assessment of the American College of Radiology Breast Imaging Reporting and Data System Atlas (5th edition), as the latter is based on descriptive categories (9). The software has been validated for DBT and SM (20,21). Continuous measures of compressed breast thickness (in millimeters), breast volume (in cubic centimeters), fibroglandular volume (in cubic centimeters, absolute dense tissue), and volumetric breast density (VBD, percentage of the breast volume) were provided by the software. VDG represents the average value for one examination from the four standard mammographic views (mediolateral oblique and craniocaudal views of each breast). We present results by VDG, quintiles of VBD, and VDG 1 and 2 versus VDG 3 and 4 (Tables E1–E3 [online]).

Recall was defined as a screening examination with mammographic findings that resulted in a recall for further assessment. SDC was defined as breast cancer (ductal carcinoma in situ or invasive breast cancer) diagnosed as a result of the recall, whereas an FP result was defined as recall for further assessment with negative outcome. Positive predictive values of recalls and biopsies were defined as the number of women diagnosed with SDC among those recalled and biopsied, respectively. The histopathologic tumor characteristics included tumor diameter, histologic grade, lymph node status, and immunohistochemical subtypes.

The unit for analyses was number of screened women. Rates of recalls, biopsies, and SDC were defined as the number of women recalled, biopsied, and diagnosed with SDC, respectively, among those screened, whereas the rate of FP was defined as the number of FP findings among the number of women screened. Histopathologic tumor characteristics were presented as percentages of women with invasive breast cancer, including invasive carcinoma or no special type ($n = 112$), invasive lobular carcinoma ($n = 19$), invasive tubular

Table 1: Baseline Characteristics of the Women Screened with DBT or Standard DM in the To-Be Trial

Characteristic	DBT+SM (*n* = 14 380)	DM (*n* = 14 369)
Age (y)*	59 (54–64)	59 (54–64)
Screening history		
Prevalent	2013 (14.0)	2053 (14.3%)
Subsequent	12 367 (86.0)	12 316 (85.7%)
Body mass index (kg/m^2)*	25 (23–28)	25 (23–28)
Without information	4378	4499
Breast volume (cm^3)*	844 (576–1171)	848 (571–1190)
Without information	121	86
Fibroglandular volume (cm^3)*	39.8 (29.6–55.0)	42.9 (32.0–58.5)
Without information	121	86
Compressed breast thickness (mm)*	60.8 (52.3–68.3)	61.0 (52.3–68.5)
Without information	121	86
Volumetric breast density (%)*	4.7 (3.2–7.6)	5.2 (3.4–8.4)
Without information	121	86
Volpara Density Grade		
1	3929 (27.6)	3212 (22.5)
2	6216 (43.6)	6280 (44.0)
3	3152 (22.1)	3655 (25.6)
4	962 (6.7)	1136 (7.8)
Without information	121	86
Volumetric breast density quintiles		
First	2804 (19.7) [1.5–3.0]	2729 (19.1%) [1.6–3.1]
Second	2843 (19.9) [3.0–4.0]	2822 (19.8%) [3.2–4.2]
Third	2844 (19.9) [4.0–5.5]	3012 (21.1%) [4.3–6.1]
Fourth	2876 (20.2) [5.6–8.6]	2830 (19.8%) [6.2–9.4]
Fifth	2892 (20.3) [8.7–40.9]	2890 (20.2) [9.5–35.6]
Without information	121	86

Note.—Unless otherwise indicated, data are numbers of women, with percentages in parentheses and the range in brackets. DBT+SM = digital breast tomosynthesis including two-dimensional synthetic mammograms, DM = digital mammography, IQR = interquartile range.

* Data are the median, and data in parentheses are the interquartile range.

carcinoma (*n* = 6), mucinous carcinoma (*n* = 5), and other types (*n* = 8).

We included data on weight and height from a questionnaire used in BreastScreen Norway from 2006 to 2016 (22). Breast volume was used as a proxy for body mass index (BMI) for women without weight and height data (23).

Statistical Analysis

The study sample was described with summary statistics, including medians with interquartile range (IQR) relative frequencies (Table 1). The differences between screening techniques were tested by comparing the 95% confidence intervals (CIs) around the proportions, presented graphically as pairwise bar graphs and by using *Z* tests. The numerical values for the graphs are shown in Table E4 (online). Differences in categorical distributions were tested using a χ^2 test.

We analyzed the relative risk (RR) of recall, FP, and SDC for DBT+SM and DM by VDG, using log-binomial regression models and adjusting for age groups (<55 years, 55–59 years, 60–64 years, and >64 years), screening history (dichotomized as prevalent and incident screens), and breast volume (continuous). The strength of breast volume as a proxy for BMI was investigated using the Pearson correlation coefficient (Table E5 [online]). We modeled the interaction between VDG and the screening technique using DM and VDG 1 as baseline categories (Table E6 [online]). Because the absolute rates of recall, FP, and SDC in VDG 1 differed for DBT+SM and DM, the RRs could not be directly compared.

A *P* value lower than .05 indicated a significant difference. All analyses *(https://github.com/anderskyrud/To-Be_density)* were performed with Stata software 16 (College Station, Tex) or *R* software (version 3.6.1; Vienna, Austria).

Results

Participant and Tumor Characteristics

Of 44 266 women invited to screening in Bergen, 32 976 attended screening and 29 453 consented to participate in the To-Be trial. These women represented the per-protocol population of the randomized controlled trial. Women were excluded if they had breast implants (*n* = 524), previous history of breast cancer (*n* = 630), or metastases from

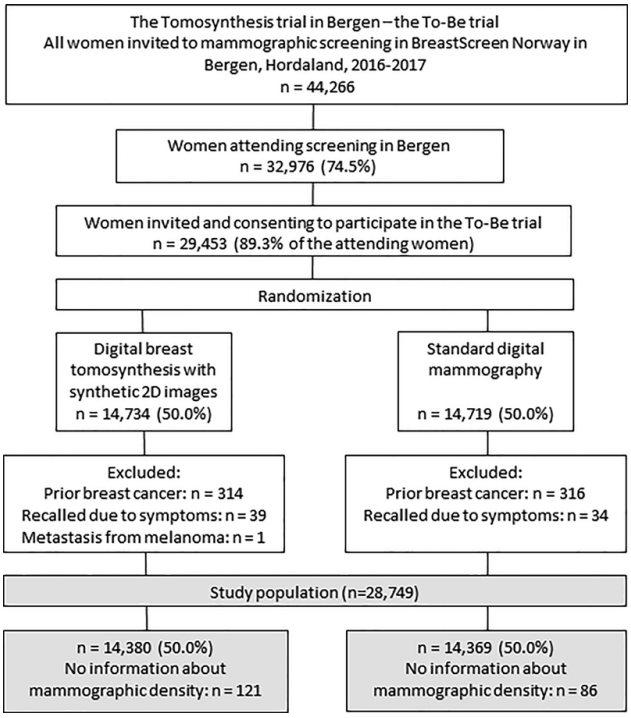

Figure 1: Flowchart shows exclusion criteria and the final study sample for women screened with digital breast tomosynthesis including synthesized two-dimensional (2D) mammography and standard digital mammography.

other cancer types (n = 1) or if they reported symptoms (n = 73). The remaining 28 749 women included 14 380 screened with DBT+SM and 14 369 screened with DM (Fig 1, Table 1). Information about VDG was missing due to random technical errors for 207 (0.7%) women, 121 in the DBT+SM arm and 86 in the DM arm, and these women were excluded from analysis (Fig 1). The images from women screened with DBT+SM and DM are shown in Figures 2 and 3. Information about BMI was available for 19 872 women, 10 002 in the DBT+SM arm, and 9870 in the DM arm.

Median age was 59 years (IQR, 54–64 years), and median BMI was 25 kg/m^2 (IQR, 23–28 kg/m^2), for women in both arms (Table 1). Median compressed breast thickness was 60.8 mm (IQR, 52.3–68.3 mm) for DBT+SM and 61.0 mm (IQR, 52.3–68.5 mm) for DM. Median breast volume was 844 cm^3 (IQR, 576–1171 cm^3) for DBT+SM and 848 cm^3 (IQR, 571–1190 cm^3) for DM. In the DBT+SM arm, median fibroglandular volume was 39.8 cm^3 (IQR, 29.6–55.0 cm^3), whereas it was 42.9 cm^3 (IQR, 32.0–58.5 cm^3) for DM. Median VBD was 4.7% (IQR, 3.2%–7.6%) for DBT+SM and 5.2% (IQR, 3.4%–8.4%) for DM. Mean VBD was lower for women in the DBT+SM arm than for those in the DM arm (6.3% ± 4.5 [standard deviation] vs 6.8% ± 4.7; $P <$ .001) (Fig E1 [online]).

The largest tumor diameter was shown for VDG 2 in the DBT+SM arm, with a mean diameter of 16.8 mm (95% CI: 13.9, 19.8), and for VDG 4 in the DM arm, with a mean diameter of 19.7 mm (95% CI: 6.0, 33.3) (Tables 2, 3). The majority of the tumors in both arms were classified as histologic grade 2 (50% [38 of 76] for DBT+SM, 50% [34 of 68] for DM) and Luminal A subtype (58.7% [44 of 75] for DBT+SM, 60.9% [42 of 69] for DM).

Recall Rates

The use of DBT+SM resulted in a lower recall rate for DBT+SM versus DM in women with VDG 1 and VDG 2, with a rate of 2.1% (81 of 3929; 95% CI: 1.6%, 2.5%) versus 3.3% (106 of 3212; 95% CI: 2.7%, 3.9%; P = .001) for VDG 1 and a rate of 3.2% (200 of 6216; 95% CI: 2.8%, 3.7%) versus 4.3% (267 of 6280; 95% CI: 3.8%, 4.8%; P = .002) for VDG 2 (Fig 4, *A*; Table E4 [online]). A difference was not detected between DBT+SM and DM for VDG 3 or 4; 4.1% (129 of 3152; 95% CI: 3.4%, 4.8%) versus 4.0% (147 of 3655; 95% CI: 3.4%, 4.7%; P = .88) for VDG 3 and 3.1% (30 of 962; 95% CI: 2.0%, 4.2%) versus 4.0% (46 of 1136; 95% CI: 2.9%, 5.2%; P = .26) for VDG 4. In the stratified analysis for DBT+SM, adjusted RR of recall was 1.8 (95% CI: 1.4, 2.4; $P <$ .001) for VDG 2, 2.4 (95% CI: 1.7, 3.3; $P <$ .001) for VDG 3, and 1.8 (95% CI: 1.1, 2.9; P = .002) for VDG 4 using VDG 1 as a reference (Table 4).

Sensitivity analyses using VBD quintiles verified increasing RRs of recall for DBT+SM by increasing density (second quintile, 1.6; 95% CI: 1.1, 2.2; P = .02; third quintile, 2.1; 95% CI: 1.4, 2.9; $P <$ .001; fourth quintile, 2.6; 95% CI: 1.8, 3.7; $P <$ .001; fifth quintile, 3.0; 95% CI: 2.1, 4.5; $P <$ .001) but not for DM (second quintile, 1.2; 95% CI: 0.9, 1.5; P = .33; third quintile, 1.3; 95% CI: 1.0, 1.8; P = .05; fourth quintile, 1.2; 95% CI: 0.9, 1.6; P = .33; fifth quintile, 1.1; 95% CI: 0.8, 1.5; P = .78) (Table E1 [online]). When combining *(a)* VDG 1 and 2 and *(b)* VDG 3 and 4, an increased RR of recall by increasing density category was observed for DBT+SM (1.4; 95% CI: 1.1, 1.7; P = .004) but not DM (0.9; 95% CI: 0.8, 1.1; P = .40) (Table E3 [online]).

FP Results

The rate of FP results was 1.6% (63 of 3929; 95% CI: 1.2%, 2.0%) versus 2.8% (91 of 3212; 95% CI: 2.3%, 3.4%; $P <$ .001) for DBT+SM versus DM for VDG 1 and 2.4% (152 of 6216; 95% CI: 2.1%, 2.8%) versus 3.6% (228 of 6280; 95% CI: 3.2%, 4.1%; $P <$ .001) for VDG 2 (Fig 4, *B*; Table E4 [online]). The rates did not differ between DBT+SM and DM for VDG 3 (3.4% [106 of 3152]; 95% CI: 2.7%, 4.0% vs 3.3% [122 of 3655]; 95% CI: 2.8%, 3.9%; P = .95) or VDG 4 (2.5% [24 of 962]; 95% CI: 1.5%, 3.5% vs 3.4% [30

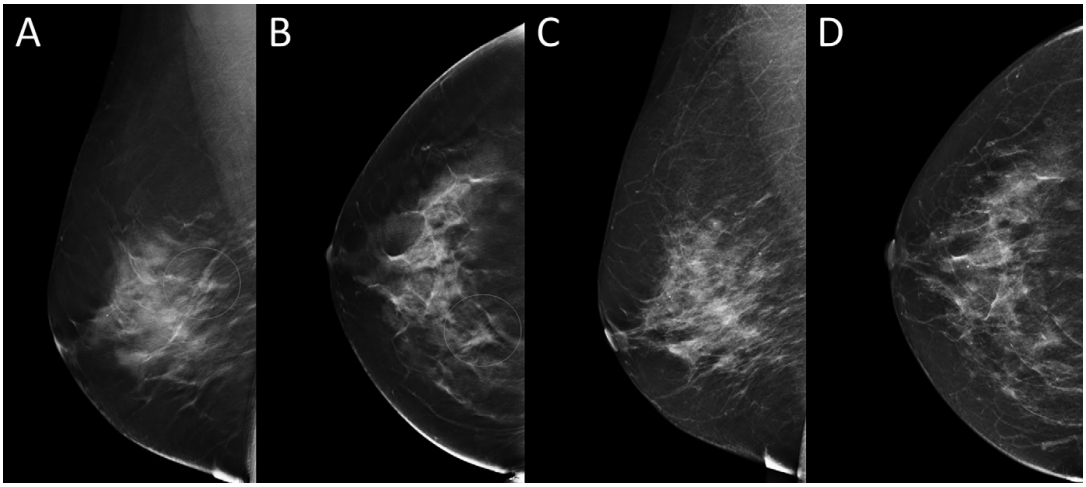

Figure 2: *A*, Right mediolateral oblique and, *B*, craniocaudal digital breast tomosynthesis images at 1-mm plane, and, *C*, right mediolateral oblique and, *D*, craniocaudal synthetic two-dimensional images in a 59-year-old woman with high breast density (Breast Imaging Reporting and Data System 3). The woman was recalled after digital breast tomosynthesis screening because of a spiculated mass only visible at 1-mm planes in both views (marked with a circle) and not visible on the synthetic two-dimensional images. Histologic examination revealed a multifocal tumor, including 12-mm invasive ductal carcinoma of no special type, grade 1–2, luminal A, and 20-mm ductal carcinoma in situ.

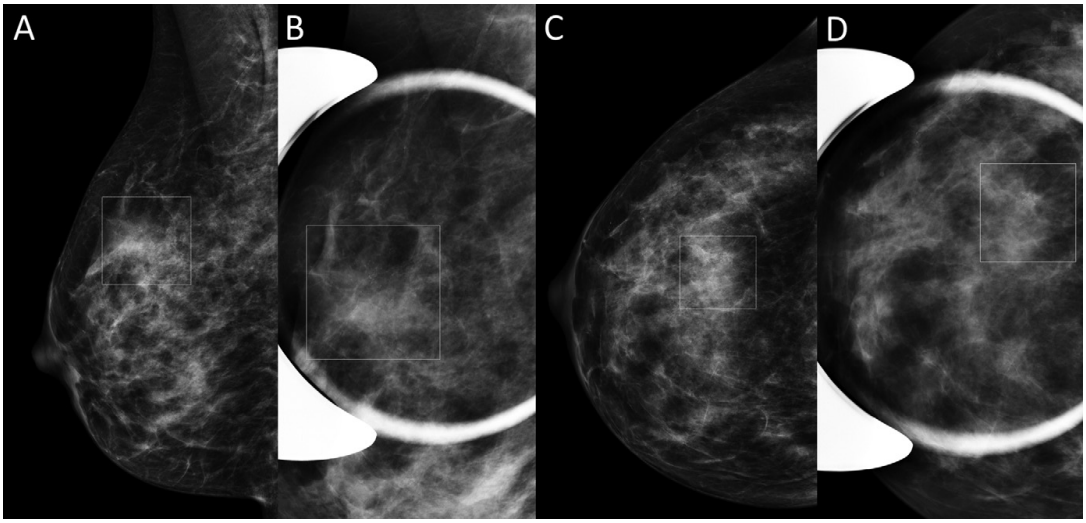

Figure 3: *A*, Right mediolateral oblique and, *C*, craniocaudal digital mammography images with, *B*, *D*, spot magnification in a 56-year-old woman with high breast density (Breast Imaging Reporting and Data System 3). Both readers detected calcifications in the central upper part of the right breast (□). Histologic examination revealed a 25-mm invasive ductal carcinoma of no special type (grade 3, luminal B).

of 1136]; 95% CI: 2.4%, 4.5%; $P = .21$). Adjusted RR of FP for DBT+SM was 1.7 (95% CI: 1.2, 2.3; $P = .001$) for VDG 2 and 2.3 (95% CI: 1.6, 3.3; $P < .001$) for VDG 3 compared with VDG 1 (Table 5).

Sensitivity analyses combining *(a)* VDG 1 and 2 and *(b)* VDG 3 and 4 showed increased RR of FP by increasing density category for DBT+SM (1.4; 95% CI: 1.1, 1.8; $P = .01$) but not for DM (0.9; 95% CI: 0.7, 1.1; $P = .15$) (Table E3 [online]).

Positive Predictive Values and SDC

Positive predictive values of recalls and biopsies were higher for DBT+SM versus DM for VDG 2 (24.0%; 1492 of 6216; 95% CI: 18.1%, 29.9% for DBT+SM vs 14.6%; 917 of 6280; 95% CI: 10.4%, 18.8% for DM; and 46.6%; 2897 of 6216; 95% CI: 37.0%, 56.2% for DBT+SM vs 30.2%; 1897 of 6280; 95% CI: 22.3%, 38.2% for DM; $P = .01$ for all; see Fig 4, *E* and *F*, and Table E4 [online]). Adjusted RR

Table 2: Distribution of Histopathologic Tumor Characteristics for Invasive Breast Cancers Diagnosed with DBT+SM in the To-Be Trial

Characteristic	Total (*n* = 80)	VDG 1 (*n* = 17)	VDG 2 (*n* = 38)	VDG 3 (*n* = 20)	VDG 4 (*n* = 5)
Tumor diameter					
Mean (mm)	16.0 (14.0, 18.0)	14.7 (9.2, 20.2)	16.8 (13.9, 19.8)	15.5 (12.5, 18.5)	15 (ND)
Without information*	11	1	3	3	4
Histologic grade (%)					
1	29.0 (19.1, 40.5)	17.7 (3.8,.43.4)	34.3 (19.1, 52.2)	36.8 (16.3, 61.6)	0 (ND)
2	50.0 (38.3, 61.7)	64.7 (38.3, 85.8)	37.1 (21.5, 55.1)	47.4 (24.5, 71.1)	100.0 (100.0, 100.0)
3	21.1 (12.5, 31.9)	17.7 (3.8, 43.4)	28.6 (14.6, 46.3)	15.8 (3.4, 39.6)	0 (ND)
Without information*	4	0	3	1	0
Lymph node status (%)					
Negative	82.3 (72.1, 90.0)	82.4 (56.6, 96.2)	86.5 (71.2, 95.5)	80.0 (56.3, 94.3)	60.0 (14.7, 94.7)
Positive	17.7 (10.0, 27.9)	17.7 (3.8, 43.4)	13.5 (4.5, 28.8)	20.0 (5.7, 43.7)	40.0 (5.3, 85.3)
Without information*	1	0	1	0	0
Subtype (%)					
Luminal A	58.7 (46.7, 69.9)	31.3 (11.0, 58.7)	69.4 (51.9, 83.7)	60.0 (36.1, 80.9)	66.7 (9.4, 99.2)
Luminal B Her2-	24.0 (14.9, 35.3)	37.5 (15.2, 64.6)	19.4 (8.2, 36.0)	25.0 (8.7, 49.1)	0 (ND)
Luminal B Her2+	6.7 (2.2, 14.9)	18.8 (4.1, 45.7)	0 (ND)	5.0 (0.1, 24.9)	33.3 (0.8, 90.6)
Her2+	4.0 (0.8, 11.3)	0 (ND)	5.6 (0.7, 18.7)	5.0 (0.1, 24.9)	0 (ND)
Triple negative	6.7 (2.2, 14.9)	12.5 (1.6, 38.4)	5.6 (0.7, 18.7)	5.0 (0.1, 24.9)	0 (ND)
Without information*	5	1	2	0	2

Note.—Characteristics in this table are distributed by Volpara Density Grade (VDG). Data in parentheses are 95% confidence intervals. DBT+SM = digital breast tomosynthesis including two-dimensional synthetic mammograms. ND = no data.

* Data are number of women.

Table 3: Distribution of Histopathologic Tumor Characteristics for Invasive Breast Cancers Diagnosed with Standard Digital Mammography in the To-Be Trial

Characteristic	Total (*n* = 70)	VDG 1 (*n* = 13)	VDG 2 (*n* = 33)	VDG 3 (*n* = 21)	VDG 4 (*n* = 3)
Tumor diameter					
Mean (mm)	14.5 (12.3, 16.8)	17.1 (11.3, 23.0)	14.6 (11.8, 17.3)	12.1 (6.7, 17.6)	19.7 (6.0, 33.3)
Without information*	11	0	7	4	0
Histologic grade (%)					
1	35.3 (24.1, 47.8)	23.1 (5.0, 53.8)	29.0 (14.2, 48.0)	52.4 (29.8, 74.3)	33.3 (0.8, 90.6)
2	50.0 (37.6, 62.4)	46.2 (19.2, 74.9)	64.5 (45.4, 80.8)	33.3 (14.6, 57.0)	33.3 (0.8, 90.6)
3	14.7 (7.3, 25.4)	30.8 (9.1, 61.4)	6.5 (0.8, 21.4)	14.3 (3.1, 36.3)	33.3 (0.8, 90.6)
Without information*	2	0	2	0	0
Lymph node status (%)					
Negative	73.9 (61.9, 83.8)	69.2 (38.6, 90.9)	72.7 (54.5, 86.7)	75.0 (50.9, 91.3)	100.0 (100.0, 100.0)
Positive	26.1 (16.3, 38.1)	30.8 (9.1, 61.4)	27.3 (13.3, 45.5)	25.0 (8.7, 49.1)	0 (ND)
Without information*	1	0	0	1	0
Subtype (%)					
Luminal A	60.9 (48.4, 72.4)	61.5 (31.6, 86.1)	66.7 (48.2, 82.0)	55.0 (31.5, 76.9)	33.3 (0.8, 90.6)
Luminal B Her2-	26.1 (16.3, 38.1)	23.1 (5.0, 53.8)	21.2 (9.0, 38.9)	35.0 (15.4, 59.2)	33.3 (0.8, 90.6)
Luminal B Her2+	10.1 (4.2, 19.8)	15.4 (1.9, 45.5)	9.1 (1.9, 24.3)	5.0 (0.1, 24.9)	33.3 (0.8, 90.6)
Her2+	1.5 (0.0, 7.8)	0 (ND)	0 (ND)	5.0 (0.1, 24.9)	0 (ND)
Triple negative	1.5 (0.0, 7.8)	0 (ND)	3.0 (0.1, 15.8)	0.0 (0.0, 0.0)	0 (ND)
Without information*	1	0	0	1	0

Note.—Characteristics in this table are distributed by Volpara Density Grade (VDG). Data in parentheses are 95% confidence intervals. ND = no data.

* Data are number of women.

of SDC increased by VDG for DBT+SM (VDG 2: 2.4; 95% CI: 1.3, 4.2; P = .004; VDG 3: 2.8; 95% CI: 1.3, 5.7; P = .01; VDG 4: 2.8; 95% CI: 1.0, 8.0; P = .05) but not for DM (VDG 2: 1.7; 95% CI: 0.9, 3.1; P = .13; VDG 3: 2.1; 95% CI: 1.0, 4.6; P = .06; VDG 4: 2.2; 95% CI: 0.8, 6.2; P = .15) (Table 6).

Sensitivity analyses using VBD quintiles verified increasing RRs of SDC in the DBT+SM models with increasing

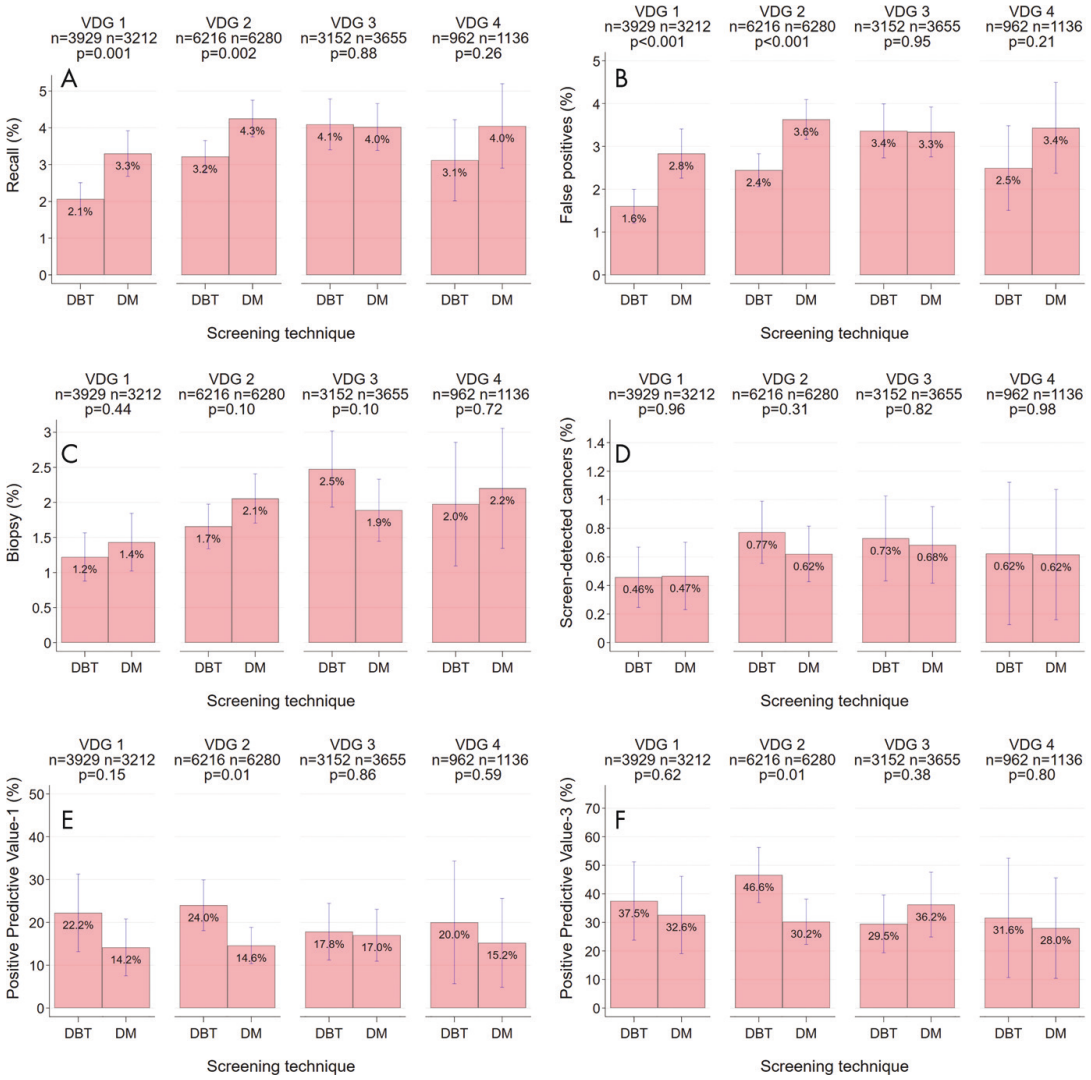

Figure 4: Rates with 95% confidence intervals of, *A*, recall, *B*, false-positive screening examinations, *C*, biopsy, *D*, screen-detected breast cancer, *E*, positive predictive value of recalls, and, *F*, positive predictive value of biopsy for digital breast tomosynthesis (DBT) that includes synthesized two-dimensional mammograms and standard digital mammography (DM) by mammographic density given as Volpara Density Grade (VDG) 1–4. Each bar represents the rate (%), whereas vertical lines represent 95% confidence intervals.

density (second quintile: 1.2; 95% CI: 0.6, 2.7; P = .61; third quintile: 2.4; 95% CI: 1.1, 5.2; P = .02; fourth quintile: 4.5; 95% CI: 2.1, 9.4; $P < .001$; fifth quintile: 3.9; 95% CI: 1.7, 9.2; $P < .001$), whereas the RR for SDC in the DM models was significant only for the third and fourth density quintile in the SDC model (second quintile: 1.6; 95% CI: 0.8, 3.6; P = .22; third quintile: 2.4; 95% CI: 1.1, 5.3; P = .03; fourth quintile: 2.8; 95% CI: 1.2, 6.4; P = .01; fifth quintile: 1.7; 95% CI: 0.7, 4.5; P = .27) (Table E1 [online]).

Discussion

As part of the To-Be trial, which randomized women to either the digital breast tomosynthesis (DBT) combined with synthetic mammography (SM) (hereafter, DBT+SM) arm or the digital mammography (DM) screening arm, the information on automated density was collected. The objective of this study was to identify differences in rates of recall, false-positive screening examinations, biopsies, and screen-detected breast cancer, and histopathologic findings for DBT+SM versus DM by automated density. Women with nondense breasts had a

Table 4: Crude and Adjusted Relative Risk of Recall for DBT+SM and Standard DM by VDG 1–4

	Relative Risk of Recall for DBT+SM						Relative Risk of Recall for DM					
Characteristic	Crude	95% CI	*P* Value	Adjusted	95% CI	*P* Value	Crude	95% CI	*P* Value	Adjusted	95% CI	*P* Value
VDG												
1	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
2	1.6	1.2, 2.0	.001	1.8	1.4, 2.4	<.001	1.3	1.0, 1.6	.03	1.3	1.0, 1.6	.06
3	2.0	1.5, 2.6	<.001	2.4	1.7, 3.3	<.001	1.2	1.0, 1.6	.11	1.1	0.8, 1.5	.41
4	1.5	1.0, 2.3	.05	1.8	1.1, 2.9	.02	1.2	0.9, 1.7	.24	1.1	0.7, 1.6	.71
Age group												
<55 years	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
55–59 years	0.5	0.4, 0.7	<.001	1.0	0.7, 1.4	.96	0.7	0.6, 0.9	.003	1.3	1.0, 1.6	.06
60–64 years	0.5	0.4, 0.7	<.001	1.0	0.7, 1.4	.99	0.7	0.6, 0.9	.002	1.1	0.8, 1.5	.41
>64 years	0.7	0.5, 0.9	.002	1.4	1.0, 1.9	.07	0.9	0.7, 1.1	.26	1.1	0.7, 1.6	.71
Screening history												
Prevalent	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
Incident	0.4	0.3, 0.5	<.001	0.4	0.3, 0.5	<.001	0.5	0.4, 0.6	<.001	0.5	0.3, 0.6	<.001
Breast volume (cm^3)	1.0	1.0, 1.0	.93	1.0	1.0, 1.0	.004	1.0	1.0, 1.0	.33	1.0	1.0, 1.0	.61

Note.—CI = confidence interval, DBT = digital breast tomosynthesis, DM = digital mammography, SM = synthetic mammography, VDG = Volpara Density Grade.

Table 5: Crude and Adjusted Relative Risk of False-Positive Findings by DBT+SM and Standard DM by VDG 1–4

	Relative Risk of FP for DBT+SM						Relative Risk of FP for DM					
Characteristic	Crude	95% CI	*P* Value	Adjusted	95% CI	*P* Value	Crude	95% CI	*P* Value	Adjusted	95% CI	*P* Value
VDG												
1	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
2	1.5	1.1, 2.0	.004	1.7	1.2, 2.3	.001	1.3	1.0, 1.6	.04	1.2	0.9, 1.6	.18
3	2.1	1.5, 2.9	<.001	2.3	1.6, 3.2	<.001	1.2	0.9, 1.5	.23	1.0	0.7, 1.4	.96
4	1.6	1.0, 2.5	.06	1.6	0.9, 2.7	.09	1.2	0.8, 1.8	.31	1.0	0.6, 1.5	.81
Age group												
<55 years	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
55–59 years	0.4	0.3, 0.6	<.001	0.8	0.5, 1.1	.15	0.7	0.5, 0.8	<.001	1.0	0.7, 1.3	.74
60–64 years	0.4	0.3, 0.6	<.001	0.8	0.5, 1.1	.14	0.6	0.5, 0.8	<.001	0.9	0.6, 1.3	.63
>64 years	0.6	0.4, 0.7	<.001	1.1	0.7, 1.5	.77	0.9	0.7, 1.1	.16	1.3	0.9, 1.7	.12
Screening history												
Prevalent	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
Incident	0.3	0.3, 0.4	<.001	0.4	0.3, 0.6	<.001	0.5	0.4, 0.6	<.001	0.5	0.4, 0.7	<.001
Breast volume (cm^3)	1.0	1.0, 1.0	.63	1.0	1.0, 1.0	.06	1.0	1.0, 1.0	.18	1.0	1.0, 1.0	.21

Note.—CI = confidence interval, DBT = digital breast tomosynthesis, DM = digital mammography, FP = false-positive, SM = synthetic mammography, VDG = Volpara Density Grade.

lower recall rate for DBT+SM compared with DM (Volpara Density Grade [VDG] 1: 2.1% [81 of 3929] vs 3.3% [106 of 3212], *P* = .001; VDG 2: 3.2% [200 of 6216] vs 4.3% [267 of 6280], *P* = .002). For women with denser breasts, the relative risk of SDC increased for DBT+SM (VDG 2: 2.4, *P* = .004; VDG 3: 2.8, *P* = .01; VDG 4: 2.8, *P* = .05) but not for DM (VDG 2: 1.7, *P* = .13; VDG 3: 2.1, *P* = .06; VDG 4: 2.2, *P* = .15). Our results support the previous conclusion that DBT is more responsive to volumetric density, where tumors, benign lesions, and normal structures are better visualized compared with DM (13,24,25). Our findings might also indicate that automated software for DBT+SM provides a tool for more discriminatory evaluation of the breast cancer risk and potential risk-stratified screening practices, as women with dense breasts might be recommended to undergo additional screening techniques and more frequent screening based on the results from DBT+SM.

Three studies have reported better performance of DBT compared with DM for recall rates and FP among women with dense breasts (2,3,16), which is in line with our findings. However, studies have shown superior accuracy of DBT compared with DM for depicting breast cancer in women with dense breasts based on Breast Imaging Reporting and Data System density categories (2,3,16,24,26). Several technical elements may have contributed to our finding of similar performance of DBT+SM in women with dense breasts compared with DM. One possible

Table 6: Crude and Adjusted RR with 95% CI of SDCs by DBT+SM and standard DM, by VDG 1–4

	RR of SDC for DBT+SM						RR of SDC for DM					
Characteristic	Crude	95% CI	*P* Value	Adjusted	95% CI	*P* Value	Crude	95% CI	*P* Value	Adjusted	95% CI	*P* Value
VDG												
1	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
2	1.7	1.0, 2.9	.06	2.34	1.3, 4.2	.004	1.3	0.7, 2.4	.35	1.7	0.9, 3.1	.13
3	1.6	0.9, 3.0	.14	2.8	1.3, 5.7	.01	1.5	0.8, 2.8	.24	2.1	1.0, 4.6	.06
4	1.4	0.5, 3.4	.51	2.8	1.0, 8.0	.05	1.3	0.5, 3.2	.54	2.2	0.8, 6.2	.15
Age group												
<55 years	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
55–59 years	1.5	0.8, 2.8	.18	3.3	1.4, 7.6	.01	1.3	0.7, 2.4	.37	3.0	1.3, 6.8	.01
60–64 years	1.5	0.8, 2.8	.18	3.6	1.5, 8.6	.004	1.4	0.8, 2.5	.29	3.2	1.4, 7.6	.01
>64 years	1.8	1.0, 3.3	.05	4.5	1.9, 10.8	.001	1.2	0.6, 2.2	.60	3.0	1.2, 7.2	.02
Screening history												
Prevalent	1.0	…	…	1.0	…	…	1.0	…	…	1.0	…	…
Incident	0.8	0.5, 1.4	.42	0.3	0.1, 0.7	.004	0.6	0.4, 1.1	.09	0.23	0.1, 0.6	.002
Breast volume (cm^3)	1.0	1.0, 1.0	.27	1.0	1.0, 1.0	.01	1.0	1.0, 1.0	.50	1.0	1.0, 1.0	.09

Note.—CI = confidence interval, DBT = digital breast tomosynthesis, DM = digital mammography, RR = relative risk, SDC = screen-detected breast cancer, SM = synthetic mammography, VDG = Volpara Density Grade.

reason is the use of SM instead of standard DM accompanying the DBT acquisition as in prior studies (27,28). Anatomic noise of structures larger than 2 mm may have limited the visibility of breast cancers in DBT in a similar manner as observed in DM (29). Moreover, DBT is known to yield better performance if tumors located in dense tissue are surrounded by some amount of fatty tissue (30). Lack of statistical differences in screening metrics for DBT+SM and DM across VDG categories and specifically among women with dense breasts might also be due to the small number of breast cancers in this trial.

Our study had several strengths. We provided outcomes from a randomized controlled trial and automated VBD, potentially eliminating inconsistencies with subjective density measurements. Sensitivity analyses with quintiles of VBD and dichotomized VDG, as well as interaction analyses, can strengthen our primary findings. The results of our study might be applied to the programs using Breast Imaging Reporting and Data System visual density assessment as well as other automated density assessment tools and DBT systems despite the discrepancies in the methods of density measurements and image acquisition (13,24,25,31,32).

The limitations of this study included the sample size and distinct population screened in Bergen, the radiologists' lack of experience with DBT+SM interpretation prior to the randomized trial, an extensive hanging protocol, first-generation DBT equipment (5,33), and missing information for VDG and tumor diameter. Our use of independent double reading with consensus is also different from usual practice in the United States. However, training according to the guidelines is required to screen-read in BreastScreen Norway (34). The lack of differences between the techniques for recall and FP rates for VDG 4 might be due to the low number of women (n = 962). High BMI is known to drive breast carcinogenesis while decreasing the relative breast density (23). We used breast volume as a proxy for BMI in our study, as we did not have information on BMI for all women (23). The median values of fibroglandular volume and VBD, as well as the proportions of women included in VDG 3 and 4, were lower for DBT+SM than for DM. The differences could be explained by discrepancies in density estimation by the software for DBT+SM and DM, which was also reported in other studies (16,28). Automated density assessment has its own limitations, including variability based on mammographic positioning (35).

In conclusion, digital breast tomosynthesis including two-dimensional synthetic mammograms (DBT+SM) was superior to digital mammography (DM) in women with lower breast density in this study. The adjusted relative risk for recall, false-positive, and screen-detected breast cancer increased by volumetric density categories for DBT+SM but not for DM. DBT+SM with automated density assessment may be a responsive and effective combination for stratified risk-based screening for breast cancer, including supplemental screening techniques or more frequent screening among women with dense breasts. More studies, combined with systematic reviews and meta-analyses, are needed to make evidence-based conclusions.

Acknowledgments: We thank the women who took part in the study, as well as the radiographers and radiologists at the breast center at Haukeland University Hospital (Bergen, Norway), for their work, support, and enthusiasm for the study.

Author contributions: Guarantors of integrity of entire study, H.S.A., S.H.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, 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, N.M., I.S.H., C.I.L., S.Z., S.H.; clinical studies, H.S.A., S.Z., S.H.; statistical analysis, N.M., A.S.D., S.H.; and manuscript editing, all authors.

Disclosures of Conflicts of Interest: N.M. disclosed no relevant relationships. **H.S.A.** disclosed no relevant relationships. **A.S.D.** disclosed no relevant relationships. **I.S.H.** disclosed no relevant relationships. **C.I.L.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: is on the GRAIL trial data safety monitoring board; institution received funding from GE Healthcare; receives textbook royalties from Wolters Kluwer, Ox-

ford University Press, and McGraw-Hill. Other relationships: disclosed no relevant relationships. **S.Z.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: received speaker's fees and travel support from Siemens Healthcare. Other relationships: holds U.S. patent no. 9,833,203. **S.H.** disclosed no relevant relationships.

References

1. Bernardi D, Gentilini MA, De Nisi M, et al. Effect of implementing digital breast tomosynthesis (DBT) instead of mammography on population screening outcomes including interval cancer rates: Results of the Trento DBT pilot evaluation. Breast 2020;50:135–140.
2. Conant EF, Barlow WE, Herschorn SD, et al. Association of digital breast tomosynthesis vs digital mammography with cancer detection and recall rates by age and breast density. JAMA Oncol 2019;5(5):635–642.
3. Zackrisson S, Lång K, Rosso A, et al. One-view breast tomosynthesis versus two-view mammography in the Malmö Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study. Lancet Oncol 2018;19(11):1493–1503.
4. Houssami N. Evidence on synthesized two-dimensional mammography versus digital mammography when using tomosynthesis (three-dimensional mammography) for population breast cancer screening. Clin Breast Cancer 2018;18(4):255–260.e1.
5. Hofvind S, Holen AS, Aase HS, et al. Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. Lancet Oncol 2019;20(6):795–805.
6. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2006;15(6):1159–1169.
7. Boyd NF, Huszti E, Melnichouk O, et al. Mammographic features associated with interval breast cancers in screening programs. Breast Cancer Res 2014;16(4):417.
8. Nazari SS, Mukherjee P. An overview of mammographic density and its association with breast cancer. Breast Cancer 2018;25(3):259–267.
9. Sickles EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS Mammography. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology, 2013.
10. Lee AY, Wisner DJ, Aminololama-Shakeri S, et al. Inter-reader variability in the use of BI-RADS descriptors for suspicious findings on diagnostic mammography: a multi-institution study of 10 academic radiologists. Acad Radiol 2017;24(1):60–66.
11. Wengert GJ, Helbich TH, Woitek R, et al. Inter- and intra-observer agreement of BI-RADS-based subjective visual estimation of amount of fibroglandular breast tissue with magnetic resonance imaging: comparison to automated quantitative assessment. Eur Radiol 2016;26(11):3917–3922.
12. Van der Waal D, den Heeten GJ, Pijnappel RM, et al. Comparing visually assessed BI-RADS breast density and automated volumetric breast density software: a cross-sectional study in a breast cancer screening setting. PLoS One 2015;10(9):e0136667.
13. Ekpo EU, McEntee MF. Measurement of breast density with digital breast tomosynthesis--a systematic review. Br J Radiol 2014;87(1043):20140460.
14. Pertuz S, McDonald ES, Weinstein SP, Conant EF, Kontos D. Fully automated quantitative estimation of volumetric breast density from digital breast tomosynthesis images: preliminary results and comparison with digital mammography and MR imaging. Radiology 2016;279(1):65–74.
15. Moshina N, Sebuødegård S, Lee CI, et al. Automated volumetric analysis of mammographic density in a screening setting: worse outcomes for women with dense breasts. Radiology 2018;288(2):343–352.
16. Østerås BH, Martinsen ACT, Gullien R, Skaane P. Digital mammography versus breast tomosynthesis: impact of breast density on diagnostic performance in population-based screening. Radiology 2019;293(1):60–68.
17. Hofvind S, Tsuruda K, Mangerud G, 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. https://www.kreftregisteret.no/globalassets/cancer-in norway/2016/mammo_cin2016_special_issue_web.pdf. Published November 9, 2017. Accessed January 15, 2020.
18. Aase HS, Holen AS, Pedersen K, 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. Eur Radiol 2019;29(3):1175–1186.
19. Aitken Z, McCormack VA, Highnam RP, et al. Screen-film mammographic density and breast cancer risk: a comparison of the volumetric standard mammogram form and the interactive threshold measurement methods. Cancer Epidemiol Biomarkers Prev 2010;19(2):418–428.
20. Machida Y, Saita A, Namba H, Fukuma E. Automated volumetric breast density estimation out of digital breast tomosynthesis data: feasibility study of a new software version. Springerplus 2016;5(1):780.
21. Conant EF, Keller BM, Pantalone L, Gastounioti A, McDonald ES, Kontos D. Agreement between breast percentage density estimations from standard-dose versus synthetic digital mammograms: results from a large screening cohort using automated measures. Radiology 2017;283(3):673–680.
22. Tsuruda KM, Sagstad S, Sebuødegård S, Hofvind S. Validity and reliability of self-reported health indicators among women attending organized mammographic screening. Scand J Public Health 2018;46(7):744–751.
23. Hudson S, Vik Hjerkind K, Vinnicombe S, et al. Adjusting for BMI in analyses of volumetric mammographic density and breast cancer risk. Breast Cancer Res 2018;20(1):156.
24. Rafferty EA, Durand MA, Conant EF, et al. Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts. JAMA 2016;315(16):1784–1786.
25. Tagliafico A, Tagliafico G, Astengo D, et al. Mammographic density estimation: one-to-one comparison of digital mammography and digital breast tomosynthesis using fully automated software. Eur Radiol 2012;22(6):1265–1270.
26. Caumo F, Zorzi M, Brunelli S, et al. Digital breast tomosynthesis with synthesized two-dimensional images versus full-field digital mammography for population screening: outcomes from the Verona screening program. Radiology 2018;287(1):37–46.
27. Østerås BH, Skaane P, Gullien R, Martinsen ACT. Average glandular dose in paired digital mammography and digital breast tomosynthesis acquisitions in a population based screening program: effects of measuring breast density, air kerma and beam quality. Phys Med Biol 2018;63(3):035006.
28. Gastounioti A, McCarthy AM, Pantalone L, Synnestvedt M, Kontos D, Conant EF. Effect of mammographic screening modality on breast density assessment: digital mammography versus digital breast tomosynthesis. Radiology 2019;291(2):320–327.
29. Chen L, Abbey CK, Nosratieh A, Lindfors KK, Boone JM. Anatomical complexity in breast parenchyma and its implications for optimal breast imaging strategies. Med Phys 2012;39(3):1435–1441.
30. García-Barquín P, Páramo M, Elizalde A, et al. The effect of the amount of peritumoral adipose tissue in the detection of additional tumors with digital breast tomosynthesis and ultrasound. Acta Radiol 2017;58(6):645–651.
31. Jeffers AM, Sieh W, Lipson JA, et al. Breast cancer risk and mammographic density assessed with semiautomated and fully automated methods and BI-RADS. Radiology 2017;282(2):348–355.
32. Zackrisson S. Tomosynthesis in breast screening: great expectations? Lancet Oncol 2019;20(6):745–746.
33. Waade GG, Holen Å, Sebuødegård S, et al. Breast compression parameters among women screened with standard digital mammography and digital breast tomosynthesis in a randomized controlled trial. Acta Radiol 2020;61(3):321–330.
34. Hofvind S, Bennett RL, Brisson J, et al. Audit feedback on reading performance of screening mammograms: An international comparison. J Med Screen 2016;23(3):150–159.
35. Tagliafico A, Tagliafico G, Houssami N. Differences in breast density assessment using mammography, tomosynthesis and MRI and their implications for practice. Br J Radiol 2013;86(1032):20130528.

Radiology

Erratum

Originally published in:
https://doi.org/10.1148/radiol.2020201150

Comparing Screening Outcomes for Digital Breast Tomosynthesis and Digital Mammography by Automated Breast Density in a Randomized Controlled Trial: Results from the To-Be Trial

Nataliia Moshina, Hildegunn S. Aase, Anders S. Danielsen, Ingfrid S. Haldorsen, Christoph I. Lee, Sophia Zackrisson, Solveig Hofvind

Erratum in:
https://doi.org/10.1148/radiol.229003

The article contained an error in the Data and Mammographic Density Measurements section. The corrected sentence should be as follows: **Histopathologic tumor characteristics were presented as percentages of women with invasive breast cancer, including invasive carcinoma or no special type (*n* = 112), invasive lobular carcinoma (*n* = 19), invasive tubular carcinoma (*n* = 6), mucinous carcinoma (*n* = 5), and other types (*n* = 8).**

10.1148/radiol.2020201150

Supplemental Material–Interaction Analysis

In this article, we present a subgroup analysis of the data from the randomized controlled trial. The main analysis found in the article consists of a set of log-binomial regression models stratified on the intervention/control variable (randomization variable), showing the relationship between selected performance indicators (outcomes) and mammographic density (exposure) within each subgroup. This is just one of several possible ways to perform this subgroup analysis. Another option could have been to stratify on the exposure variable instead of the randomization variable, creating density subgroups. A third option is to model a factorial interaction between the randomization variable and the exposure variable. This analytical option is provided for the reader in Table E6 to show how different analytical approaches would affect the conclusions of the article. We opted for stratification instead of interaction because an interaction model can both be difficult to interpret, and can lead to an excessive amount of terms, thus consuming degrees of freedom. However, an interaction term modeling approach allows for a more direct comparison of the relative risks. Our interaction term model showed a statistically significant increase in RR in DBT*VDG3 for both recall and FP (1.6, 95% CI: 1.1–2.3, $P = .01$; and 1.8, 95% CI: 1.2–2.7, $P = .01$), supporting the main conclusion that mammographic density is associated with recalls and false positives in the DBT group, but not in the DM group.

Table E1: Crude and adjusted Relative Risk (RR) of recall and screen-detected breast cancer (SDC) with 95% confidence interval (95% CI), by screening technique (digital breast tomosynthesis including synthetic 2D mammograms, DBT, or standard digital mammography, DM) and quintiles of volumetric breast density

Volumetric breast density	RR of recall for DBT						RR of recall for DM					
	Crude	95% CI	*P* value	Adjusted*	95% CI	*P* value	Crude	95% CI	*P* value	Adjusted*	95% CI	*P* value
first quintile	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
second quintile	1.3	0.9–1.9	0.12	1.6	1.1–2.2	0.02	1.2	0.9–1.6	0.23	1.2	0.9–1.5	0.33
third quintile	1.6	1.1–2.2	0.01	2.1	1.4–2.9	<0.001	1.4	1.1–1.8	0.02	1.3	1.0–1.8	0.05
fourth quintile	1.9	1.4–2.7	<0.001	2.6	1.8–3.7	<0.001	1.2	1.0–1.6	0.12	1.2	0.9–1.6	0.33
fifth quintile	2.2	1.6–3.0	<0.001	3.0	2.1–4.5	<0.001	1.2	0.9–1.6	0.22	1.1	0.8–1.5	0.78
	RR of SDC for DBT						RR of SDC for DM					
first quintile	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
second quintile	0.9	0.4–2.0	0.81	1.2	0.6–2.7	0.61	1.4	0.7–3.0	0.38	1.6	0.78–3.6	0.22
third quintile	1.4	0.7–2.9	0.31	2.4	1.1–5.2	0.02	1.9	0.9–3.9	0.08	2.4	1.1–5.3	0.03
fourth quintile	2.3	1.2–4.3	0.01	4.5	2.1–9.4	<0.001	2.0	1.0–4.1	0.06	2.8	1.2–6.4	0.01

fifth quintile	1.6	0.8–3.1	0.20	3.9	1.7–9.2	<0.001	1.1	0.5–2.5	0.79	1.7	0.7–4.5	0.27

* Adjusted for age groups (< 55; 55–59; 60–64; and > 64 years), screening history (prevalent or incident) and breast volume (cm^3).

Table E2: Crude and adjusted Relative Risk (RR) of recall and screen-detected breast cancer (SDC) with 95% confidence interval (95% CI), by screening technique (digital breast tomosynthesis including synthetic 2D mammograms, DBT, or standard digital mammography, DM) and quintiles of mammographic density. Sensitivity analysis using body mass index in the adjustment of mammographic density–value available for 69.1% of the women/observations

Volumetric breast density	RR of recall for DBT						RR of recall for DM					
	Crude	95% CI	*P* value	Adjusted*	95% CI	*P* value	Crude	95% CI	*P* value	Adjusted*	95% CI	*P* value
first quintile	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
second quintile	1.7	1.1–2.6	0.03	1.6	1.1–2.2	0.02	1.2	0.9–1.6	0.23	1.1	0.8–1.6	0.47
third quintile	1.8	1.2–2.9	0.008	2.1	1.4–2.9	<0.001	1.4	1.1–1.8	0.02	1.5	1.1–2.0	0.02
fourth quintile	2.6	1.7–4.0	<0.001	2.6	1.8–3.7	<0.001	1.2	1.0–1.6	0.12	1.2	0.9–1.8	0.24
fifth quintile	2.8	1.8–4.4	<0.001	3.0	2.1–4.5	<0.001	1.2	0.9–1.6	0.22	1.0	0.7–1.5	0.85
	RR of SDC for DBT						RR of SDC for DM					
first quintile	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
second quintile	0.9	0.4–2.0	0.81	1.0	0.4–2.4	0.92	1.4	0.7–3.0	0.38	1.9	0.7–5.1	0.21
third quintile	1.4	0.7–2.9	0.31	1.5	0.7–3.5	0.34	1.9	0.9–3.9	0.08	2.6	1.0–6.7	0.06
fourth quintile	2.3	1.2–4.3	0.01	2.4	1.1–5.4	0.03	2.0	1.0–4.1	0.06	2.8	1.0–7.5	0.04
fifth quintile	1.6	0.8–3.1	0.20	2.3	0.9–5.5	0.07	1.1	0.5–2.5	0.79	1.0	0.3–3.5	0.94

*Adjusted for age groups (< 55; 55–59; 60–64; and > 64 years), screening history (prevalent or incident) and body mass index (kg/m^2).

Table E3: Crude and adjusted Relative Risk (RR) of recall, false positive (FP) screening examinations, and screen-detected breast cancer (SDC) by screening technique (digital breast tomosynthesis including synthetic 2D mammograms, DBT, or standard digital mammography, DM) and Volpara Density Grade (VDG)

	RR of recall for DBT						RR of recall for DM					
	Crude	95% CI	*P* value	Adjusted*	95% CI	*P* value	Crude	95% CI	*P* value	Adjusted*	95% CI	*P* value
VDG 1+2	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
VDG 3+4	1.4	1.1–1.7	0.001	1.4	1.1–1.7	0.004	1.0	0.9–1.2	0.78	0.9	0.8–1.1	0.40
	RR of FP for DBT						RR of FP for DM					
VDG 1+2	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
VDG 3+4	1.5	1.2–1.9	<0.001	1.4	1.1–1.8	0.01	1.0	0.8–1.2	0.99	0.9	0.7–1.1	0.15
	RR of SDC for DBT						RR of SDC for DM					

VDG 1+2	1.0	—	—	1.0	—	—	1.0	—	—	1.0	—	—
VDG 3+4	1.1	0.7–1.7	0.72	1.4	0.8–2.2	0.24	1.2	0.8–1.8	0.47	1.4	0.8–2.3	0.21

*Adjusted for age groups (< 55; 55–59; 60–64; and > 64 years), screening history (prevalent or incident) and breast volume (cm^3).

Table E4: Crude numbers and point estimates with 95% confidence interval (95% CI), for screening metrics by screening technique (digital breast tomosynthesis including synthetic 2D mammograms, DBT, or standard digital mammography, DM), and Volpara density grade (VDG) as shown in Figures 2A–2F

	DBT			
Screening metrics	VDG 1	VDG 2	VDG 3	VDG 4
	n = 3929	*n* = 6216	*n* = 3152	*n* = 962
Consensus	4.1% (3.5%–4.8%)	6.7% (6.1%–7.3%)	7.8% (6.9%–8.8%)	7.5% (5.8%–9.1%)
False positive examinations	1.6% (1.2%–2.0%)	2.4% (2.1%–2.8%)	3.4% (2.7%–4.0%)	2.5% (1.5%–3.5%)
Recall	2.1% (1.6%–2.5%)	3.2% (2.8%–3.7%)	4.1% (3.4%–4.8%)	3.1% (2.0%–4.2%)
Biopsies	1.2% (0.9%–1.6%)	1.7% (1.3%–2.0%)	2.5% (1.9%–3.0%)	2.0% (1.1%–2.9%)
Screen-detected cancer	0.46% (0.25%–0.67%)	0.77% (0.55%–0.99%)	0.73% (0.43%–1.03%)	0.62% (0.13%–1.12%)
PPV-1	22.2% (13.2%–31.3%)	24.0% (18.1%–29.9%)	17.8% (11.2%–24.4%)	20.0% (5.7%–34.3%)
PPV-3	37.5% (23.8%–51.2%)	46.6% (37.0%–56.2%)	29.5% (19.4%–39.6%)	31.6% (10.7%–52.5%)
	DM			
	VDG 1	VDG 2	VDG 3	VDG 4
	n = 3212	*n* = 6280	*n* = 3655	*n* = 1136
Consensus	5.9% (5.1%–6.7%)	7.8% (7.1%–8.4%)	7.9% (6.9%–8.7%)	7.9% (6.4%–9.5%)
False positive examinations	2.8% (2.3%–3.4%)	3.6% (3.2%–4.1%)	3.3% (2.8%–3.9%)	3.4% (2.4%–4.5%)
Recall	3.3% (2.7%–3.9%)	4.3% (3.8%–4.8%)	4.0% (3.4%–4.7%)	4.0% (2.9%–5.2%)
Biopsies	1.4% (1.0%–1.8%)	2.1% (1.7%–2.4%)	1.9% (1.4%–2.3%)	2.2% (1.3%–3.1%)
Screen-detected cancer	0.47% (0.23%–0.70%)	0.62% (0.43%–0.82%)	0.68% (0.42%–0.95%)	0.61% (0.16%–1.07%)
PPV-1	14.2% (7.5%–20.8%)	14.6% (10.4%–18.8%)	17.0% (10.9%–23.1%)	15.2% (4.8%–25.6%)
PPV-3	32.6% (19.1%–46.2%)	30.2% (22.3%–38.2%)	36.2% (24.9%–47.6%)	28.0% (10.4%–45.6%)

Table E5: Correlation matrix for BMI and automated measured breast volume*

	Body mass index	Breast volume (cm^3)	Volumetric breast density (%)	Fibroglandular volume (cm^3)
Body mass index (kg/m^2)	1			
Breast volume (cm^3)	0.62	1		
Volumetric breast density (%)	–0.41	–0.54	1	
Fibroglandular volume (cm^3)	0.07	0.23	0.53	1

* All correlations were significant at < 0.001.

Table E6: Crude and adjusted Relative Risk (RR) of recall, false positives (FP) and screen-detected breast cancer (SDC)

	RR of recall					
	Crude	95% CI	*P* value	Adjusted[#]	95% CI	*P* value
Volpara Density grade						
VDG 1	1.0	—	—	1.0	—	—
VDG 2	1.4	1.2–1.7	<0.001	1.4	1.1–1.7	0.01
VDG 3	1.6	1.3–1.9	<0.001	1.3	1.0–1.7	0.08
VDG 4	1.4	1.1–1.8	<0.001	1.3	0.9–1.8	0.25
Screening technique						
DM	1.0	—	—	1.0	—	—
DBT	0.8	0.7–0.9	<0.001	0.6	0.5–0.8	0.002
Screening technique * Volpara Density Grade						
VDG1*DM	—	—	—	1.0	—	—
VDG1*DBT	—	—	—	1.0	—	—
VDG2*DM	—	—	—	1.0	—	—
VDG2*DBT	—	—	—	1.2	0.9–1.7	0.26
VDG3*DM	—	—	—	1.0	—	—
VDG3*DBT	—	—	—	1.6	1.1–2.3	0.01
VDG4*DM	—	—	—	1.0	—	—
VDG4*DBT	—	—	—	1.2	0.7–2.1	0.49
Age groups						
<55 years	1.0	—	—	1.0	—	—
55–59 years	0.6	0.5–0.8	<0.001	1.1	0.9–1.3	0.61
60–64 years	0.6	0.5–0.7	<0.001	1.1	0.8–1.3	0.67
>64 years	0.8	0.6–0.9	0.004	1.4	1.1–1.7	0.003
Screening history						
Prevalent	1.0	—	—	1.0	—	—
Incident	0.5	0.4–0.5	<0.001	0.4	0.4–0.5	<0.001
Breast volume, cm^3	1.0	1.0–1.0	0.51	1.0	1.0–1.0	0.14
	RR of FP for DBT					
	Crude	95% CI	*P* value	Adjusted[#]	95% CI	*P* value
Volpara Density grade						

VDG 1	1.0	—	—	1.0	—	—
VDG 2	1.4	1.2–1.7	<0.001	1.3	1.0–1.6	0.06
VDG	1.6	1.3–1.9	<0.001	1.1	0.8–1.5	0.47
VDG 4	1.4	1.0–1.9	0.03	1.1	0.7–1.6	0.75
Screening technique						
DM	1.0	—	—	1.0	—	—
DBT	0.7	0.6–0.8	<0.001	0.6	0.4–0.8	<0.001
Screening technique * Volpara Density Grade						
VDG1*DM	—	—	—	1.0	—	—
VDG1*DBT	—	—	—	1.0	—	—
VDG2*DM	—	—	—	1.0	—	—
VDG2*DBT	—	—	—	1.2	0.8–1.7	0.36
VDG3*DM	—	—	—	1.0	—	—
VDG3*DBT	—	—	—	1.8	1.2–2.7	0.01
VDG4*DM	—	—	—	1.0	—	—
VDG4*DBT	—	—	—	1.3	0.7–2.3	0.43
Age groups						
<55 years	1.0	—	—	1.0	—	—
55–59 years	0.5	0.5–0.7	<0.001	0.9	0.9–1.1	0.23
60–64 years	0.5	0.4–0.6	<0.001	0.9	0.7–1.1	0.18
>64 years	0.7	0.6–0.9	<0.001	1.2	0.9–1.5	0.17
Screening history						
Prevalent	1.0	—	—	1.0	—	—
Incident	0.4	0.4–0.5	<0.001	0.5	0.4–0.6	<0.001
Breast volume, cm^3	1.0	1.0–1.0	0.19	1.0	1.0–1.0	0.84
	RR of SDC for DBT					
	Crude	95% CI	*P* value	Adjusted[#]	95% CI	*P* value
Volpara Density grade						
VDG 1	1.0	—	—	1.0	—	—
VDG 2	1.5	1.0–2.3	0.04	1.8	0.9–3.2	0.08
VDG 3	1.5	1.0–2.4	0.06	2.4	1.2–4.8	0.02
VDG 4	1.3	0.7–2.5	0.37	2.5	1.0–6.6	0.07
Screening technique						
DM	1.0	—	—	1.0	—	—
DBT	1.1	0.8–1.5	0.56	1.0	0.5–2.0	0.97

Screening technique * Volpara Density Grade						
VDG1*DM	—	—	—	1.0	—	—
VDG1*DBT	—	—	—	1.0	—	—
VDG2*DM	—	—	—	1.0	—	—
VDG2*DBT	—	—	—	1.3	0.6–2.9	0.55
VDG3*DM	—	—	—	1.0	—	—
VDG3*DBT	—	—	—	1.1	0.4–2.6	0.89
VDG4*DM	—	—	—	1.0	—	—
VDG4*DBT	—	—	—	1.0	0.3–3.6	0.98
Age groups						
<55 years	1.0	—	—	1.0	—	—
55–59 years	1.4	0.9–2.2	0.12	3.1	1.8–5.6	<0.001
60–64 years	1.6	0.9–2.2	0.09	3.4	1.8–6.3	<0.001
>64 years	1.5	1.0–2.3	0.07	3.7	2.0–6.9	<0.001
Screening history						
Prevalent	1.0	—	—	1.0	—	—
Incident	0.7	0.5–1.0	0.08	0.3	0.2–0.5	<0.001
Breast volume, cm^3	1.0	1.0–1.0	0.21	1.0	1.0–1.0	0.002

Adjusted for age groups (< 55; 55–59; 60–64; and > 64 years), screening history (prevalent or incident) and breast volume (cm^3).

VDG = Volpara density grade, DBT = digital breast tomosynthesis including synthetic 2D mammograms, DM = standard digital mammography.

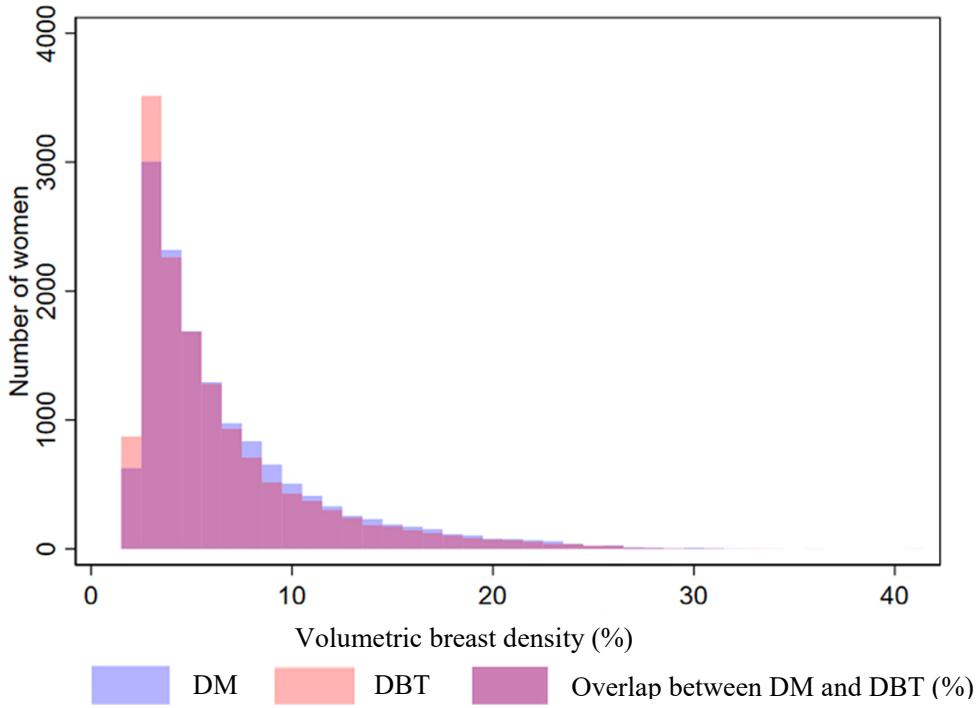

Figure E1: Automated measures of volumetric breast density (VBD) for women randomly selected for screening with digital breast tomosynthesis including synthetic 2D mammograms (DBT) or standard digital mammography (DM).

Paper 3

European Journal of Radiology 141 (2021) 109753

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

Mammographic features and screening outcome in a randomized controlled trial comparing digital breast tomosynthesis and digital mammography

HS Aase [a,b], AS Danielsen [c,d], SR Hoff [e], ÅS Holen [c], IS Haldorsen [a,b,f], T Hovda [g], B Hanestad [a], CK Sandvik [a], S Hofvind [c,h,*]

[a] Department of Radiology, Haukeland University Hospital, Bergen, Norway
[b] Department of Clinical Medicine, University of Bergen, 5020, Bergen, Norway
[c] Section for Breast Cancer Screening, Cancer Registry of Norway, Oslo, Norway
[d] Norwegian Institute of Public Health, Oslo, Norway
[e] Department of Radiology, Møre and Romsdal Hospital Trust, Ålesund, Norway
[f] Centre for Medical Imaging and Visualization, Haukeland University Hospital, Bergen, Norway
[g] Department of Radiology, Vestre Viken Hospital Trust, Drammen, Norway
[h] Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

ARTICLE INFO

Keywords:
Randomized controlled trial
Breast cancer
Mass screening
Digital breast tomosynthesis
Mammography

ABSTRACT

Purpose: To compare the distribution of mammographic features among women recalled for further assessment after screening with digital breast tomosynthesis (DBT) versus digital mammography (DM), and to assess associations between features and final outcome of the screening, including immunohistochemical subtypes of the tumour.

Methods: This randomized controlled trial was performed in Bergen, Norway, and included 28,749 women, of which 1015 were recalled due to mammographic findings. Mammographic features were classified according to a modified BI-RADS-scale. The distribution were compared using 95 % confidence intervals (CI).

Results: Asymmetry was the most common feature of all recalls, 24.3 % (108/444) for DBT and 38.9 % (222/571) for DM. Spiculated mass was most common for breast cancer after screening with DBT (36.8 %, 35/95, 95 %CI: 27.2–47.4) while calcifications (23.0 %, 20/87, 95 %CI: 14.6–33.2) was the most frequent after DM. Among women screened with DBT, 0.13 % (95 %CI: 0.08–0.21) had benign outcome after recall due to indistinct mass while the percentage was 0.28 % (95 %CI: 0.20–0.38) for DM. The distributions were 0.70 % (95 %CI: 0.57–0.85) versus 1.46 % (95 %CI: 1.27–1.67) for asymmetry and 0.24 % (95 %CI: 0.16–0.33) versus 0.54 % (95 %CI: 0.43–0.68) for obscured mass, among women screened with DBT versus DM, respectively. Spiculated mass was the most common feature among women diagnosed with non-luminal A-like cancer after DBT and after DM.

Conclusions: Spiculated mass was the dominant feature for breast cancer among women screened with DBT while calcifications was the most frequent feature for DM. Further studies exploring the clinical relevance of mammographic features visible particularly on DBT are warranted.

1. Introduction

Mammography is the most common screening tool for breast cancer. During the last decades, standard digital mammography (DM) has replaced screen-film mammography in the Western part of the world [1, 2]. However, digital breast tomosynthesis (DBT) is expected to be the future screening tool for breast cancer [3–5]. European studies have reported higher rates of screen-detected breast cancer when comparing

Abbreviations: DBT, digital breast tomosynthesis including synthetic 2D mammography; DM, standard digital mammography; IC, interval cancer; PPV, positive predictive value for recalls; SDC, screen detected cancers; SM, synthetic two-dimensional mammography.

* Corresponding author at: Cancer Registry of Norway, P.O. 5313 Majorstuen, 0304, Oslo, Norway.
E-mail addresses: hildegunn.aase@helse-bergen.no (H. Aase), AndersSkyrud.Danielsen@fhi.no (A. Danielsen), Solveig.Roth.Hoff@helse-mr.no (S. Hoff), asne.holen@kreftregisteret.no (Å. Holen), ingfrid.helene.salvesen.haldorsen@helse-bergen.no (I. Haldorsen), Tone.Hovda@vestreviken.no (T. Hovda), berit.hanestad@helse-bergen.no (B. Hanestad), Cecilia.karolina.sandvik@helse-bergen.no (C. Sandvik), solveig.hofvind@kreftregisteret.no (S. Hofvind).

https://doi.org/10.1016/j.ejrad.2021.109753
Received 3 February 2021; Received in revised form 17 April 2021; Accepted 30 April 2021
Available online 5 May 2021

DBT alone or in combination with DM/synthetic mammograms (SM) versus standard DM [2,6–10]. Recall rates seem to vary in prospective studies [4,8,9,11]. Higher rate of screen-detected breast cancer in DBT is expected to reduce the number of interval cancers although the few published studies on interval cancer have lacked statistical power to conclude on this [12–16]. An increased rate of screen-detected cancer, without a simultaneous reduction in interval cancer rate, might indicate that DBT detects small and biologically less aggressive cancers, potentially representing small, low proliferation tumors, which could represent overdiagnosis and thus cause overtreatment [13,14].

The Tomosynthesis trial in Bergen (the To-Be trial), a randomized controlled trial conducted in Bergen, Norway, compared screening outcome for DBT + SM versus DM [17]. The rates of screen-detected breast cancer did not differ statistically for the two techniques, thus not reproducing results from other studies showing a substantial higher rate of screen-detected cancer among those screened with DBT [4,8,9,11,18,19]. However, the To-Be trial showed that recall rate and rate of false positive screening examinations were lower for DBT than for DM [17]. The somewhat unexpected results from the To-Be trial might relate to the use of first generation equipment, limited experiences in screen-reading DBT among the breast radiologists, or the perception and/or interpretation of mammographic features.

The specific mammographic features that lead to recalls and the diagnosis of screen-detected breast cancer are well documented for DM, and their correlation with histopathological characteristics have been reported [20–23]. Spiculated masses are more often estrogen- and progesterone receptor positive, HER2 negative and with lower proliferative activity compared with other masses, all indicating less aggressive tumors. Calcifications in general are associated with ductal carcinoma in situ (DCIS) and with invasive ductal carcinoma in combination with DCIS, while casting calcifications are associated with non-luminal-cancer, more often histologic grade 3 and decreased overall survival [20–23]. Less is known about mammographic features among women recalled after DBT and most of the published studies have focused on mammographic features of breast cancer while features with benign outcome is less investigated [9,11,24–27].

Overlapping breast tissue might resemble mammographic abnormalities in DM, thereby causing false positive screening results, or the opposite, the overlapping tissue might obscure tumors, resulting in false negative screening results [11,28]. DBT is known to reduce the effect of overlapping breast tissue, thereby improving visualization of both malignant and benign findings [28]. Better understanding of the features, and their association with malignant versus benign/negative outcome is thus warranted.

To gain knowledge about mammographic features of women recalled for further assessment after screening with DBT + SM (hereafter referred as DBT) versus standard DM, we analyzed data collected as a part of the To-Be trial. This study aimed to compare the distribution of mammographic features in women recalled after screening with DBT versus DM, and assess associations between features and final outcome of the screening examination, including immunohistochemically subtypes of the tumours. Our hypothesis was that recalls due to masses would result in a higher percentage of breast cancer for women screened with DBT versus DM.

2. Material and methods

The To-Be trial was a randomized controlled trial approved by the Regional Committees for Medical and Health Research Ethics (2015/424) and registered at ClinicalTrials.gov (NCT02835625). Written informed consent from all participating women was obtained. The trial was conducted in Bergen, as a part of BreastScreen Norway during one screening round, in 2016 and 2017. BreastScreen Norway is a population-based screening program for breast cancer, administered by the Cancer Registry of Norway. The program invites women aged 50–69 years to two-view mammography biennially. The screening program and the trial is described in detail elsewhere [17,29–31].

2.1. Study sample

A total of 28,749 women were included in the To-Be trial; 14,380 screened with DBT and 14,369 with DM [17]. Among those screened with DBT, 444 (3.1 %) were recalled due to mammographic findings while the corresponding number for DM was 571 (4.0 %). The recalled women comprised the study sample in this study (Fig. 1). We received a pseudonymized dataset from the Cancer Registry of Norway, containing information about the women's screening examination and recall assessment. Data included diagnostic procedures, mammographic features and histopathological findings. The DBT arm included 95 breast malignancies; 80 invasive cancers and 15 ductal carcinoma in situ (DCIS) whereas the DM arm included 87 malignancies; 71 invasive cancers and 16 DCIS (Figs. 2–4). Invasive cancers with DCIS components was considered invasive.

2.2. Screen-reading and consensus

All women underwent two-view (cranio-caudal and medio-lateral oblique) DBT or DM of both breasts. We used first generation equipment from GE (Senographe Essential SenoClaire 3D Breast Tomosynthesis™) for imaging. Eight radiologists with varying experience in breast radiology and screen-reading (0–20 years) participated in the screen-reading [29]. All screening mammograms were independently read by two breast radiologists. The hanging protocol included two sets of prior screening mammograms, with even older images available at the workstation, (GE Healthcare MammoWorkstation Version 4.7.0 Image Diagnost). Mammograms with suspicious findings indicated by one or both radiologists (n = 1968) were discussed in a consensus meeting, including two or more radiologists, where 48 % of the cases were dismissed, leaving 1015 women recalled for further assessment, 444 for DBT and 571 for DM (Fig. 1).

2.3. Recall assessment

Recall assessments were performed by the same eight radiologists who did the screen-reading. Recalled women underwent additional imaging (ultrasound alone or in combination with DM and/or DBT) and clinical examination before the radiologist decided whether a biopsy was needed. The diagnostic biopsies were performed under ultrasound or stereotactic guidance. MRI was performed in women with lobular cancer confirmed with needle biopsy, highly suspicious findings in combination with mammographic dense breast (Breast Imaging Reporting and Data System, BI-RADS, c or d [32]), and when neoadjuvant treatment was considered, according to national guidelines in Norway [33]. We used contrast-enhanced spectral mammography in women with suspicious MRI-findings without an ultrasound-correlate, and in women with contraindications for MRI (pacemakers or claustrophobia).

2.4. Variables of interest

We reported mean age at screening (years) and screening history for the recalled women. Screening history was defined as prevalent (first screening examination in BreastScreen Norway) or subsequent screening examination.

Recall was defined as further assessment due to mammographic findings. The outcome of the recall could be positive or negative. Positive was defined as ductal carcinoma in situ or invasive breast cancer, hereafter referred as breast cancer, while negative was defined as no cancer diagnosed after additional imaging alone or in combination with a needle biopsy. Positive predictive value of the recalls (PPV-1) was defined as breast cancer diagnosed among the women recalled. Positive predictive value of biopsies (PPV-3) was defined as breast cancer diagnosed among those biopsied.

At consensus, before the women were recalled, the radiologists

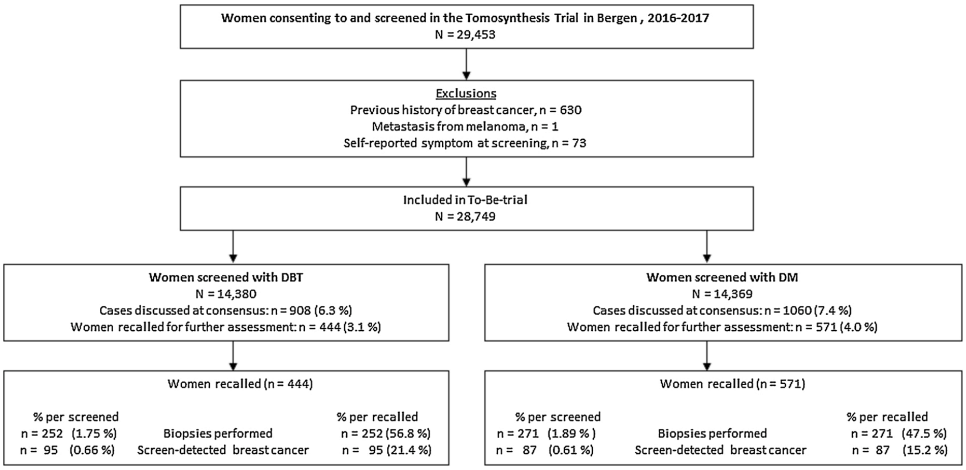

Fig. 1. Number (n) and percentage (%) of women included in the To-Be trial 2016–2017, discussed at consensus and recalled for assessment due to mammographic findings, biopsies performed and breast cancer detected, by screening technique.

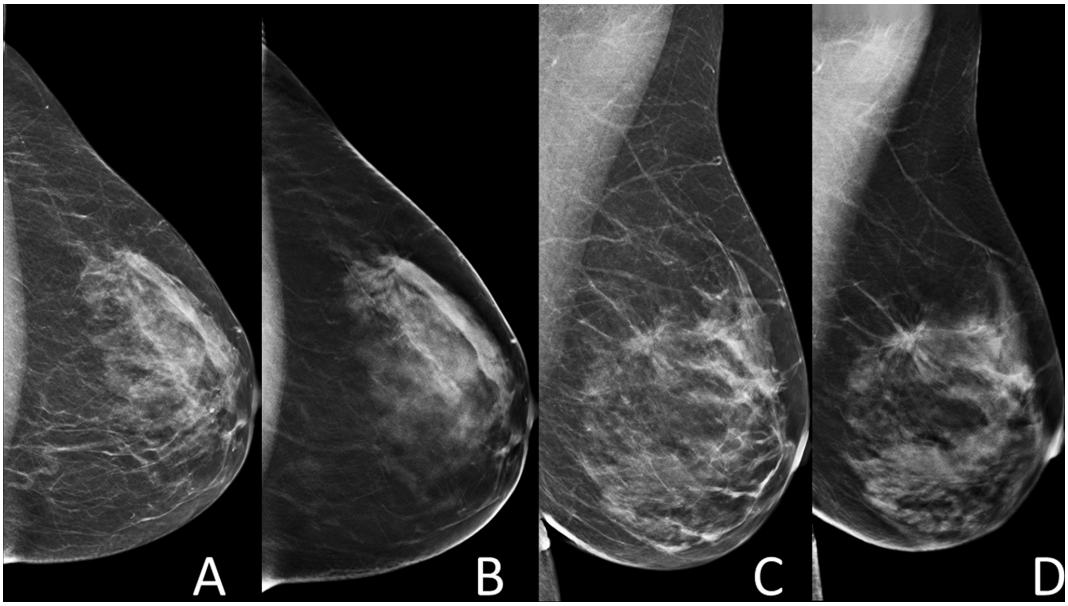

Fig. 2. Left craniocaudal synthetic 2D image (A) and 1 mm plane (B), left mediolateral oblique synthetic 2D image (C) and 1 mm plane (D), in a woman recalled after DBT because of spiculated mass in the lateral upper part of left breast. Histologic examination revealed a nonluminal A-like invasive carsinoma, histologic grade 1 and ductal carsinoma in situ grade 2.

classified the women's mammographic density according to BI-RADS and mammographic features to a modified BI-RADS-scale [30,32]. Circumscribed mass was defined as a mass with more than 75 % of the margin being well-defined and no part of the margin appearing indistinct. Obscured mass was defined as a mass with less than 75 % of the margin being well-defined and no part of the margin appearing indistinct. The category indistinct mass was used when the whole or parts of the margin was indistinct (poorly defined) or microlobulated as defined in BI-RADS [32]. We defined a mass including calcifications "mass with calcifications" while spiculated mass, architectural distortion, asymmetry, calcifications, and associated features were defined according to BI-RADS [32].

Invasive cancers were histologically classified into five subtypes based on immunohistochemistry [34] and collapsed into two groups; luminal A-like and non-luminal A-like (Luminal B HER2-, Luminal B HER2+, HER2+, and triple negative. Low Ki67-level was defined as Ki-67 level <30 %, high level as Ki-67 $\geq$ 30 %).

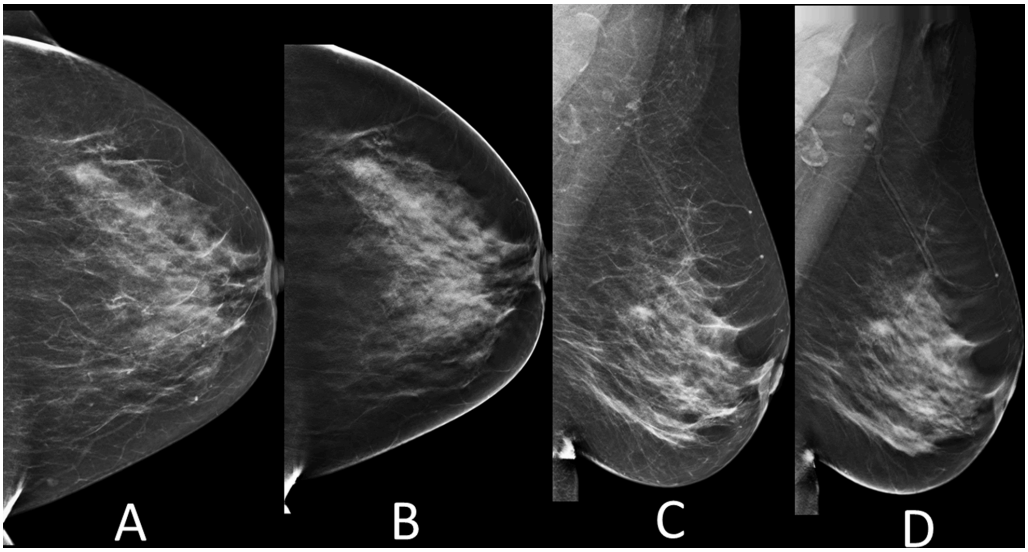

Fig. 3. Both readers picked this indistinct tumor in the lateral part of the left breast in this woman screened with DBT. **A and B:** Left craniocaudal synthetic 2D image and 1 mm plane, **C and D:** Left mediolateral oblique synthetic 2D image and 1 mm plane. The tumor measured 18 mm at histology, and was a nonluminal A-like invasive carcinoma NST, histologic grade 3.

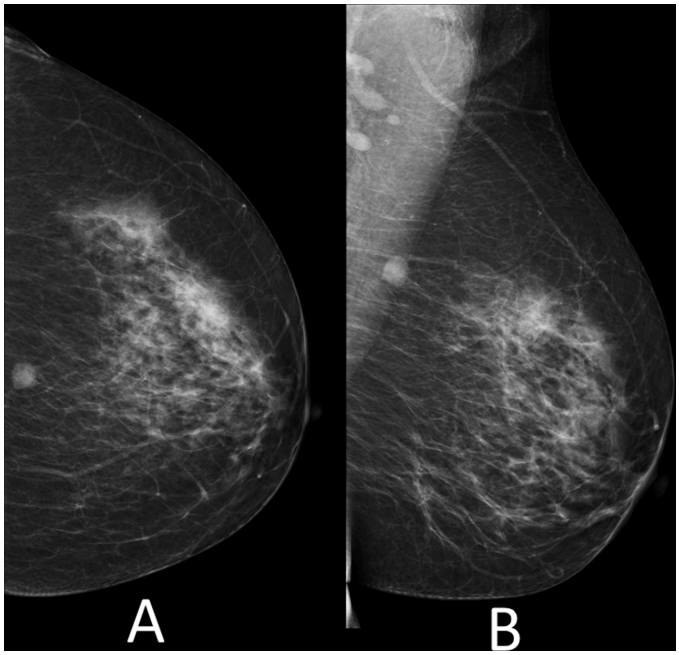

Fig. 4. Left craniocaudal (A) and left mediolateral oblique (B) image in woman recalled after screening with digital mammography. This was a 13 mm luminal A-like invasive carcinoma NST, histologic grade 1.

2.5. Statistical analyses

Descriptive results were presented for DBT and DM separately. Mean and standard deviations of the age for recalled women was described in years. Screening history, mammographic density, assessment method used at recall, use of needle biopsy and mammographic features were presented as numbers and percentages with 95 % confidence intervals (CI) among the recalled women.

Number and percentage of mammographic features were presented with different denominators; a) positive and negative recall assessment; b) mammographic feature; and c) number of screened women, for DBT and for DM. The distribution of mammographic features for luminal A-like or non-luminal A-like cancers were presented as number and percentage by a) subgroups and b) mammographic feature, for DBT and DM. We tested for differences between the two screening techniques using 95 % CI. The statistical package Stata (version 15; Texas, USA) was used for all data analyses.

3. Results

Age, screening history and mammographic density did not differ statistically for women recalled after screening with DBT (n = 444) versus DM (n = 571) (Table 1). All recalled women irrespective of screening technique underwent ultrasound as a part of their assessments, whereas 85.6 % (380/444) and 89.0 % (508/571) of those recalled after DBT and DM, respectively, had other imaging modalities in addition to ultrasound. A higher proportion of the women recalled after DBT 56.8 % (252/444, 95 %CI: 52.0–61.4) had a needle biopsy compared to those screened and recalled after DM, 47.5 % 271/571 (95 %CI: 43.3–51.6). PPV-1 was 21.4 % for DBT versus 15.2 % for DM, while PPV-3 was 37.7 % and 32.1 %, respectively.

The most common mammographic feature among the recalled women was asymmetry, 24.3 % for DBT (108/444, 95 %CI: 20.4–28.6) and 38.9 % for DM (222/571, 95 %CI: 34.9–43.0) (Table 2). A higher percentage of women were recalled due to circumscribed mass and architectural distortions after screening with DBT compared to DM; for circumscribed mass; 18.7 % (83/444, 95 %CI: 15.2–22.6) versus 8.1 % (46/571, 95 %CI:6.0–10.6) and architectural distortion; 15.8 % (70/444, 95 %CI: 12.5–19.5) versus 7.9 %, (45/571, 95 %CI:5.8–10.49) for DBT and DM, respectively. An obscured mass was less frequently observed after DBT (7.9 %, 35/444, 95 %CI: 5.6–10.8) compared with DM (14.2 %, 81/571, 95 %CI: 11.4–17.3).

Among the recalled women with positive outcome/breast cancer 36.8 % (35/95, 95 %CI: 27.2–47.4) cases diagnosed after screening with DBT were classified as spiculated mass while it was 18.4 % (16/87, 95 %CI: 10.9–28.1) for DM (Table 3a). Indistinct mass was the second most frequent feature among the cancer cases both for DBT, 16.8 % (16/95, 95 %CI: 9.9–25.9) and DM, 18.4 % (16/87, 95 %CI: 10.9–28.1). Calcifications was observed in 13.7 %, (13/95, 95 %CI: 7.5–22.3) of the cancer cases for DBT and 23.0 % (20/87, 95 %CI: 14.6–33.2) for DM. Among the recalled cases with negative outcome, asymmetry (Figs. 5 and 6) and obscured mass were less common features in DBT compared to DM. Asymmetry was found in 28.9 % (101/349, 95 %CI: 24.2–34.0) of the negative cases after recall screening with DBT versus 43.4 % (210/484, 95 %CI:38.9–47.9) after DM, and obscured mass in 9.7 % (34/349, 95 %CI: 6.8–13.3) after DBT versus 16.1 % (78/484, 95 %CI: 13.0–19.7) after DM.

Among women recalled due to asymmetry, negative outcome was observed in 93.5 % (101/108, 95 %CI: 87.1–97.4) for those screened with DBT and 94.6 % (210/222, 95 %CI: 90.7–97.2) for DM (Table 3b). Negative outcome after recall for indistinct mass was observed in 54.3 % (19/35, 95 %CI: 36.6–71.2) for DBT and 71.4 % (40/56, 95 %CI: 57.8–82.7) for DM.

Using the number of screened women in the denominator, the percentage of breast cancer classified as spiculated mass was 0.24 % (35/14,380, 95 %CI: 0.17–0.34) for DBT compared to 0.11 % (16/14,369, 95 %CI: 0.06–0.18) for DM (Table 3c). The percentage of benign outcome was 0.13 % (19/14380, 95 %CI: 0.08–0.21) for indistinct mass among women screened with DBT versus 0.28 % (40/14369, 95 %CI: 0.20–0.38 for DM, asymmetry 0.70 % (101/14380, 95 %CI:0.57–0.85) versus 1.46 % (210/14369, 95 %CI: 1.27–1.67) and obscured mass 0.24 % (34/14380, 95 %CI:0.16–0.33) versus 0.54 % (74/14369, 95 %CI: 0.43–0.68).

Among women diagnosed with invasive breast cancer after screening with DBT, 58.7 % (44/75, 95 %CI: 46.7–69.9) were luminal A-like compared to 61.4 % (43/70, 95 %CI: 49.0–72.8) of the women screened with DM (Table 4). For DBT, 52.3 % (23/44, 95 %CI: 36.7–67.5) of the luminal A-like cancers were classified as spiculated mass compared to 20.9 % (9/43, 95 %CI: 10.0–36.0) after DM. Spiculated mass was the most frequent feature among non-luminal A-like cancers, 29.0 % (9/31, 95 %CI: 14.2–48.0) after screening with DBT and 25.9 % (7/27, 95 %CI: 11.1–46.3) after screening with DM. Among malignant indistinct masses, 53.3 % (8/15, 95 %CI: 26.6–78.7) were non-luminal A-like for DBT versus 31.3 % (5/16, 95 %CI: 11.0–58.7) for DM.

Table 1
Characteristics of women recalled for assessment due to mammographic findings, methods used in the assessment, numbers (n) and percentages (%) of women who had a needle biopsy, positive predictive value of recalls (PPV-1) and of performed biopsies (PPV-3) after screening with digital breast tomosynthesis (DBT) or digital mammography (DM), in the To-Be trial.

	DBT (N = 444)		**DM** (N = 571)	
	n	% (95 % CI)	n	% (95 % CI)
Age (mean, standard deviation)	444	58.5 (6.4)	571	59.0 (6.3)
Screening history				
Prevalent screens (n, %)	128	28.8 (24.7–33.3)	136	23.8 (20.4–27.5)
Subsequent screens (n, %)	316	71.2 (66.7–75.3)	435	76.2 (72.5–79.6)
Mammographic density				
BIRADS a (n, %)	17	3.8 (2.2–6.1)	21	3.7 (2.3–5.6)
BIRADS b (n, %)	281	63.3 (58.6–67.8)	348	60.9 (56.8–65.0)
BIRADS c (n, %)	137	30.9 (26.6–35.4)	193	33.8 (29.9–37.8)
BIRADS d (n, %)	9	2.0 (0.9–3.8)	9	1.6 (0.7–3.0)
Assessment				
Ultrasound alone	64	14.9 (11.3–18.0)	63	11.0 (8.6–13.9)
Ultrasound and other imaging[a]	380	85.6 (82.0–88.7)	508	89.0 (86.1–91.4)
Biopsy (n, %)	252	56.8 (52.0–61.4)	271	47.5 (43.3–51.6)
PPV-1	95/444	21.4 (17.7–25.5)	87/571	15.2 (12.4–18.5)
PPV-3	95/252	37.7 (31.7–44.0)	87/271	32.1 (26.6–38.0)

[a] DM, DBT, contrast enhanced spectral mammography and/or Magnetic resonance imaging.

Table 2
Distribution (n and %) of mammographic features in women recalled for assessment due to mammgraphic findings, by screening technique (digital breast tomosynthesis, DBT, or digital mammography, DM), in the To-Be trial, 2016-2017.

	DBT (N = 444)		**DM (N = 571)**	
Mammographic features	n	% (95 % CI)	n	% (95 % CI)
Asymmetry	108	24.3 (20.4–28.6)	222	38.9 (34.9–43.0)
Circumscribed mass	83	18.7 (15.2–22.6)	46	8.1 (6.0–10.6)
Architectural distortion	70	15.8 (12.5–19.5)	45	7.9 (5.8–10.4)
Calcifications	49	11.0 (8.3–14.3)	78	13.7 (10.9–16.8)
Spiculated mass	37	8.3 (5.9–11.3)	16	2.8 (1.6–4.5)
Indistinct mass	35	7.9 (5.6–10.8)	56	9.8 (7.5–12.5)
Obscured mass	35	7.9 (5.6–10.8)	81	14.2 (11.4–17.3)
Mass with calcifications	22	5.0 (3.1–7.4)	21	3.7 (2.3–5.6)
Associated features	4	0.9 (0.2–2.3)	5	0.9 (0.3–2.0)
No information	1	0.2 (0.0–1.2)	1	0.2 (0.0–1.0)

Table 3
Number and distribution (n,%) of mammographic features for recalled women with positive (invasive breast cancer and/or ductal carcinoma in situ) and negative (benign after assessment with or without needle biopsy) outcome by a) recall outcome, b) by mammographic features, c) by rates of screened women, stratified by screening technique (digital breast tomosynthesis, DBT and digital mammography, DM) in the To-Be trial, 2016-2017.

a) By recall outcome	DBT (N = 444)				DM (N = 571)			
	Positive n = 95		Negative n = 349		Positive n = 87		Negative n = 484	
	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)
Spiculated mass	35	36.8 (27.2–47.4)	2	0.6 (0.1–2.1)	16	18.4 (10.9–28.1)	–	
Indistinct mass	16	16.8 (9.9–25.9)	19	5.4 (3.3–8.4)	16	18.4 (10.9–28.1)	40	8.3(6.0–11.1)
Calcifications	13	13.7 (7.5–22.3)	36	10.3 (7.3–14.0)	20	23.0 (14.6–33.2)	58	12.0 (9.2–15.2)
Architectural distortion	10	10.5 (5.2–18.5)	60	17.2 (13.4–21.6)	7	8.0 (3.3–15.9)	38	7.9 (5.6–10.6)
Asymmetry	7	7.4 (3.0–14.6)	101	28.9 (24.2–34.0)	12	13.8 (7.3–22.9)	210	43.4 (38.9–47.9)
Mass with calcifications	7	7.4 (3.0–14.6)	15	4.3 (2.4–7.0)	11	12.6 (6.5–21.5)	10	2.1 (1.0–3.8)
Circumscribed mass	5	5.3 (1.7–11.9)	78	22.3 (18.1–27.1)	2	2.3 (0.3–8.1)	44	9.1 (6.7–12.0)
Obscured mass	1	1.1 (0.0–5.7)	34	9.7 (6.8–13.3)	3	3.4 (0.7–9.7)	78	16.1 (13.0–19.7)
Associated features	1	1.1 (0.0–5.7)	3	0.9 (0.2–2.5)	–		5	1.0 (0.3–2.4)
No information	–		1	0.3 (0.0–1.6)	–		1	0.2 (0.0–1.1)
b) By mammographic feature	DBT (N = 444)				DM (N = 571)			
	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)
Spiculated mass	35	94.6 (81.8–99.3)	2	5.4 (0.7–18.2)	16	100 (79.4–1)	–	
Indistinct mass	16	45.7 (28.8–63.4)	19	54.3 (36.6–71.2)	16	28.6 (17.3–42.2)	40	71.4 (57.8–82.7)
Calcifications	13	26.5 (14.9–41.1)	36	73.5 (58.9–85.1)	20	25.6 (16.4–36.8)	58	74.4 (63.2–83.6)
Architectural distortion	10	14.3 (7.1-–24.7)	60	85.7 (75.3–92.9)	7	15.6 (70.5–93.5)	38	84.4(70.5–93.5)
Asymmetry	7	6.5 (2.6–12.9)	101	93.5 (87.1–97.4)	12	5.4 (2.8–9.3)	210	94.6 (90.7–97.2)
Mass with calcifications	7	31.8 (13.9–54.9)	15	68.2 (45.1–86.1)	11	52.4 (29.8–74.3)	10	47.6(25.7–70.2)
Circumscribed mass	5	6.0 (2.0–13.5)	78	94.0 (81.9–95.7)	2	4.3 (0.5–14.8)	44	95.7 (85.2–99.5)
Obscured mass	1	2.9 (0.1–14.9)	34	97.1 (85.1–99.9)	3	3.7 (0.8–10.4)	78	96.3 (89.6–99.2)
Associated features	1	25.0 (0.6–80.6)	3	75.0 (19.4–99.4)	–		5	100 (47.8–1)
No information			1	100 (2.5–1)	–		1	100 (2.5–1)
c) By screened women	DBT (N = 14,380)				DM (N = 14,369)			
	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)
Spiculated mass	35	0.24 (0.17–0.34)	2	0.01 (0.00–0.05)	16	0.11 (0.06–0.18)	–	
Indistinct mass	16	0.11 (0.06–0.18)	19	0.13 (0.08–0.21)	16	0.11 (0.06–0.18)	40	0.28 (0.20–0.38)
Calcifications	13	0.09 (0.05–0.15)	36	0.25 (0.18–0.35)	20	0.14 (0.09–0.21)	58	0.40 (0.31–0.52)
Architectural distortion	10	0.07 (0.03–0.13)	60	0.42 (0.32–0.54)	7	0.05 (0.02–0.10)	38	0.26 (0.19–0.36)
Asymmetry	7	0.05 (0.02–0.10)	101	0.70 (0.57–0.85)	12	0.08 (0.04–0.15)	210	1.46 (1.27–1.67)
Mass with calcifications	7	0.05 (0.02–0.10)	15	0.10 (0.06–0.17)	11	0.08 (0.04–0.14)	10	0.07 (0.03–0.13)
Circumscribed mass	5	0.03 (0.01–0.08)	78	0.54 (0.43–0.68)	2	0.01 (0.00–0.05)	44	0.31 (0.22–0.41)
Obscured mass	1	0.01 (0.00–0.04)	34	0.24 (0.16–0.33)	3	0.02 (0.00–0.06)	78	0.54 (0.43–0.68)
Associated features	1	0.01 (0.00–0.04)	3	0.02 (0.00–0.06)	–		5	0.03 (0.01–0.08)
No information	–		1	0.01 (0.00–0.04)	–		1	0.01 (0.00–0.04)

4. Discussion

In this study, we observed differences in the distribution of mammographic features for women recalled after screening with DBT versus DM. Asymmetry was the most common feature of all recalls for DBT and for DM, although less frequent for DBT compared to DM. Spiculated mass was the most common feature among women recalled and diagnosed with breast cancer after screening with DBT, while calcification was most frequent for recalled women diagnosed with breast cancer after screening with DM. Further, spiculated mass was the most common feature among women diagnosed with a non-luminal A-like cancer after DBT and after DM. The percentage of asymmetries, indistinct and obscured masses in women with a negative outcome after recall was lower for DBT versus DM.

Our finding of spiculated mass being the most common mammographic feature (36.8 %) for cancers detected after screening with DBT is in line with other studies. The Oslo Tomosynthesis Screening Trial showed a comparable rate (37 %) [25], while it was 68 % in the Malmö Breast Tomosynthesis Screening Trial [11]. The higher percentage in the Malmö trial might be due to use of different classification systems; To-Be 2 used five categories of masses, the Oslo study three (circumscribed, mass with calcifications and spiculated), while the Malmö-trial used two; circumscribed and spiculated. The distribution of immunohistochemical subtypes did not differ for DBT versus DM in our study, which was in line with results from the Malmö-trial [35]. Some studies have reported that spiculated masses are associated with less aggressive luminal A-like cancers [7,36,37]. However, both in our and the Malmö study, spiculated masses were the most common mammographic feature among the non-luminal A-like cancers, after DBT as well as after DM [35].

Indistinct mass might be easier to classify "correctly" with DBT compared to DM because the thin planes visualize tumor margins more clearly than DM. In the Malmö-trial, circumscribed mass was the second most common non-luminal-A-like cancer, which again differ from our results probably due to their limited number of feature-categories. Our study indicated that indistinct mass is an important feature for detecting cancers; it was the second most common feature among the breast cancers and about half of these cases were non-luminal A-like after screening with DBT.

In the To-Be trail, a low percentage (7.4 %) of the cancers detected at DBT was classified as asymmetry, which correspond to results reported from Spain (1% (1/92)) [9] and from the Oslo Tomosynthesis Trial (4% (4/101)) [25]. This finding supports the notion that overlapping tissue is less of a challenge in DBT compared to DM; soft tissue lesions are frequently visible in both views and correctly classified as a mass rather than asymmetry if real. This might indicate that use of DBT has the

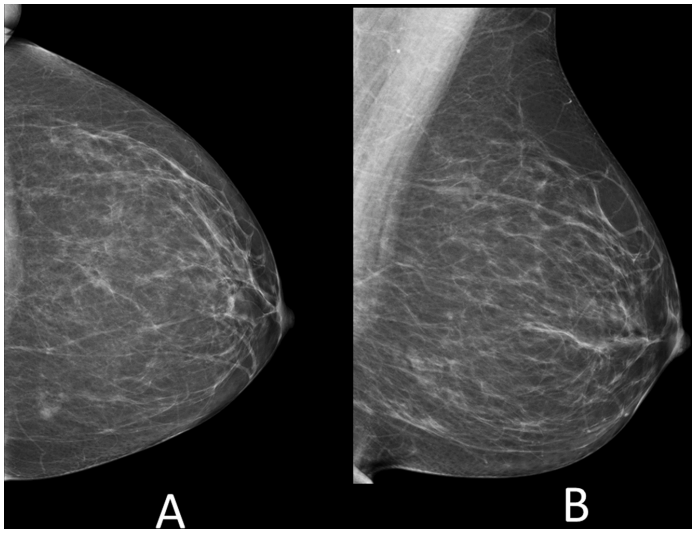

Fig. 5. Left craniocaudal (A) and left mediolaterale oblique image (B) in a woman recalled because of asymmetry in the medial part of the craniocaudal image. Assessment was performed with negative outcome, without biopsy.

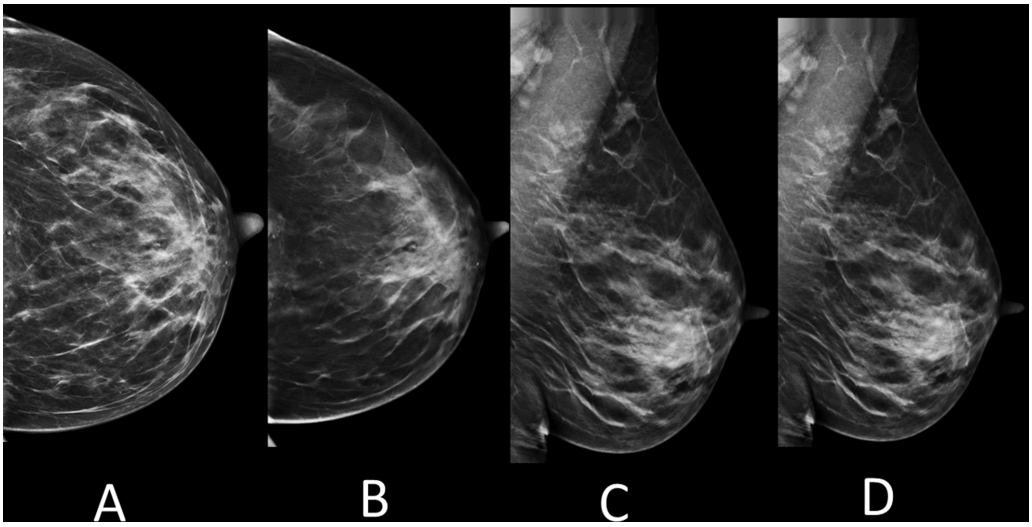

Fig. 6. Left craniocaudal synthetic 2D image (A) and 1 mm plane (B), left mediolateral oblique synthetic 2D image (C) and 1 mm plane (D), in a woman recalled after DBT, because of asymmetry in the upper part of left breast. Additional imaging at recall revealed no malignancy.

potential to reduce recalls due to asymmetry.

Calcifications were the most common feature for breast cancers detected by DM, statistically not different compared to DBT. In DBT, images are reconstructed from raw-data and calcifications are enhanced and visible, but the characterization of the calcifications might be different in DBT compared to DM [38]. The use of first-generation equipment may also have influenced our results; optimized versions of equipment are now available, which is said to visualize calcifications differently compared to first generation.

Distortion of normal architecture is a part of spiculated masses and architectural distortions [26,32]. Our results support the notion that architectural distortions are better visualized in DBT versus DM [26,39]. However distortions did not reveal a higher rate of breast cancers for DBT compared to DM. In other studies [19,24,40], higher proportions of tumors with favorable characteristics were observed for screening with DBT compared with DM. In our study, the proportion of luminal A-like cancers, a subtype known to be associated with a more favorable prognosis, did not differ between DBT and DM. This may be explained by the use of prior mammograms in the screen-reading; in the To-Be-trial priors up to 10 years back in time were available. If similar distortions or densities were identified on priors, the findings were often dismissed either at the screen-reading or at consensus. Hanging protocols are

Table 4
Distribution (n and %) of mammographic features for subgroups by screening technique (digital breast tomosynthesis, DBT, and digital mammography, DM) among invasive breast cancer cases in each subgroup (a) and among the specific features (b) in the To-Be trial, 2016-2017.

	DBT (N = 75)				DM (N = 70)			
	Luminal A-like		Non-luminal A-like		Luminal A-like		Non-luminal A-like	
	n = 44 (58.7 %)		n = 31 (41.3 %)		n = 43 (61.4 %)		n = 27 (38.6 %)	
	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)	n	% (95 % CI)
a) By subgroups								
Spiculated mass	23	52.3 (36.7–67.5)	9	29.0 (14.2–48.0)	9	20.9 (10.0–36.0)	7	25.9 (11.1–46.3)
Indistinct mass	7	15.9 (6.6–30.1)	8	25.8 (11.9–44.6)	11	25.6 (13.5–41.2)	5	18.5 (6.3–38.1)
Circumscribed mass	2	4.5 (0.6–15.5)	3	9.7 (2.0–25.8)	1	2.3 (0.1–12.3)	–	
Architectural distortion	7	15.9 (6.6–30.1)	3	9.7 (2.0–25.8)	2	4.7 (0.6–15.8)	3	11.1 (2.4–29.2)
Mass with calcifications	2	4.5 (0.6–15.5)	3	9.7 (2.0–25.8)	4	9.3 (2.6–22.1)	4	14.8 (4.2–33.7)
Asymmetry	3	6.8 (1.4–18.7)	2	6.5 (0.8–21.4)	8	18.6 (8.4–33.4)	3	11.1 (2.4–29.2)
Calcifications	–		2	6.5 (0.8–21.4)	7	16.3 (6.8–30.7)	4	14.8 (4.2–33.7)
Associated features	–		1	3.2 (0.1–16.7)	–		–	
Obscured mass	–		–		1	2.3 (0.1–12.3)	1	3.7 (0.1–19.0)
b) By mammographic feature								
Spiculated mass	23	71.9 (53.3–86.3)	9	28.1 (13.7–46.7	9	56.3 (29.9–80.2)	7	43.8 (19.8–70.1)
Indistinct mass	7	46.7 (21.3–73.4)	8	53.3 (26.6–78.7)	11	68.8 (41.3–89.0)	5	31.3 (11.0–58.7)
Circumscribed mass	2	40.0 (5.3–85.3)	3	60.0 (14.7–94.7)	1	100 (2.5–1)	–	
Architectural distortion	7	70.0 (34.8–93.3)	3	30.0 (6.7–65.2)	2	40.0 (5.3–85.3)	3	60.0 (14.7–94.7)
Mass with calcifications	2	40.0 (5.3–85.3)	3	60.0 (14.7–94.7)	4	50.0 (15.7–84.3)	4	50.0 (15.7–84.3)
Asymmetry	3	60.0 (14.7–94.7)	2	40.0 (5.3–85.3)	8	72.7 (39.0–94.0)	3	27.3 (39.0–94.0)
Calcifications	–		2	100 (15.8–1)	7	63.6 (30.8–89.1)	4	37.4 (10.9–69.2)
Associated features	–		1	100 (2.5–1)	–	–	–	–
Obscured mass	–		–	–	1	50.0 (2.5–1)	1	50.0 (2.5–1)

Information not available for 6 cases (5 DBT, 1 DM) due to neoadjuvant treatment.

usually based on "expert opinion" as evidence based guidelines are not available. Research aimed at identifying efficient hanging protocols is therefore desired.

Our finding of a higher proportion of circumscribed mass among women recalled after DBT (19 %) versus DM (8%) was unexpected since circumscribed mass is usually considered benign, not warranting a recall [32]. Lack of experience in DBT-screening among the screen-readers in the To-Be trial might explain this finding. DBT usually visualizes circumscribed mass clearly while overlapping tissue partially or totally can mask the same lesion when using DM. Notably, even though circumscribed mass represented fewer cancers compared to other features, it still contributed to 9.7 % of non-luminal A-like cancers for DBT which is in line with established knowledge; some aggressive triple negative cancers may present as indistinct-, obscure or, circumscribed masses [23].

This study, based on data from a randomized controlled trial has several limitations. The distribution of mammographic features cannot be directly compared with results from other studies due to use of different classification systems and equipment. Further, the number of cases within each mammographic feature is small and the distribution might be influenced by the absence of higher cancer detection rate for DBT versus DM in our study, contrary to other studies from Europe [2, 6–10]. A review of prior mammograms of interval and consecutive round screen-detected cancer according to features is planned, but delayed due to the covid pandemic. Limited experience in screen-reading DBT among the radiologists and use of first generation equipment from GE might also be of influence of the consensus, recall and detection rates. Further, the To-Be trial was a single center study, in which the generalizability of results should be interpreted with care.

In conclusion, this study identified different distributions of mammographic features among women recalled after screening with DBT or DM in the To-Be trial. Asymmetry was the most common feature of all recalls, however less frequent for DBT versus DM. Spiculated mass was the dominant feature for breast cancer among women screened with DBT while calcifications was the most frequent feature for DM. Further studies exploring the clinical relevance of the different mammographic features are warranted; more knowledge might enable radiologists to improve the benefit-harm-ratio in screening.

Funding

This study was funded by the Research Council of Norway (the To-Be trial; project number 247941/H10), the Cancer Registry of Norway, and the Department of Radiology at Haukeland University Hospital.

Declaration of Competing Interest

All the authors declare no conflict of interest.

Acknowledgements

We would like to thank the participating women, radiographers, radiologists and others at the breast centre, as well as the head at the Department of Radiology at Haukeland University Hospital for their support and enthusiasm for the study.

References

[1] E.D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J.K. Baum, S. Acharyya, E. F. Conant, L.L. Fajardo, L. Bassett, C. D'Orsi, R. Jong, M. Rebner, Diagnostic performance of digital versus film mammography for breast-cancer screening, N. Engl. J. Med. 353 (17) (2005) 1773–1783.

[2] P. Skaane, A.I. Bandos, R. Gullien, E.B. Eben, U. Ekseth, U. Haakenaasen, M. Izadi, I.N. Jebsen, G. Jahr, M. Krager, S. Hofvind, Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration, Eur. Radiol. 23 (8) (2013) 2061–2071.

[3] M. Alabousi, N. Zha, J.P. Salameh, L. Samoilov, A.D. Sharifabadi, A. Pozdnyakov, B. Sadeghirad, V. Freitas, M.D.F. McInnes, A. Alabousi, Digital breast tomosynthesis for breast cancer detection: a diagnostic test accuracy systematic review and meta-analysis, Eur. Radiol. 30 (4) (2020) 2058–2071.

[4] M.L. Marinovich, K.E. Hunter, P. Macaskill, N. Houssami, Breast Cancer Screening using tomosynthesis or mammography: a meta-analysis of Cancer detection and recall, J. Natl. Cancer Inst. 110 (9) (2018) 942–949.

[5] S.J. Yun, C.W. Ryu, S.J. Rhee, J.K. Ryu, J.Y. Oh, Benefit of adding digital breast tomosynthesis to digital mammography for breast cancer screening focused on cancer characteristics: a meta-analysis, Breast Cancer Res. Treat. 164 (3) (2017) 557–569.

[6] D. Bernardi, P. Macaskill, M. Pellegrini, M. Valentini, C. Fanto, L. Ostillio, P. Tuttobene, A. Luparia, N. Houssami, 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, Lancet Oncol. 17 (8) (2016) 1105–1113.

[7] K. Lang, I. Andersson, A. Rosso, A. Tingberg, P. Timberg, S. Zackrisson, 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, Eur. Radiol. 26 (1) (2016) 184–190.
[8] P. Pattacini, A. Nitrosi, P. Giorgi Rossi, V. Iotti, V. Ginocchi, S. Ravaioli, R. Vacondio, L. Braglia, S. Cavuto, C. Campari, Digital mammography versus digital mammography plus tomosynthesis for breast Cancer screening: the reggio Emilia tomosynthesis randomized trial, Radiology 288 (2) (2018) 375–385.
[9] S. Romero Martín, J.L. Raya Povedano, M. Cara García, A.L. Santos Romero, M. Pedrosa Garriguet, M. Álvarez Benito, Prospective study aiming to compare 2D mammography and tomosynthesis + synthesized mammography in terms of cancer detection and recall. From double reading of 2D mammography to single reading of tomosynthesis, Eur. Radiol. 28 (6) (2018) 2484–2491.
[10] P. Skaane, A.I. Bandos, E.B. Eben, I.N. Jebsen, M. Krager, U. Haakenaasen, U. Ekseth, M. Izadi, S. Hofvind, R. Gullien, Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images, Radiology 271 (3) (2014) 655–663.
[11] S. Zackrisson, K. Lang, A. Rosso, K. Johnson, M. Dustler, D. Fornvik, H. Fornvik, H. Sartor, P. Timberg, A. Tingberg, I. Andersson, One-view breast tomosynthesis versus two-view mammography in the Malmo Breast Tomosynthesis Screening Trial (MBTST): a prospective, population-based, diagnostic accuracy study, Lancet Oncol. 19 (11) (2018) 1493–1503.
[12] D. Bernardi, M.A. Gentilini, M. De Nisi, M. Pellegrini, C. Fantò, M. Valentini, V. Sabatino, A. Luparia, N. Houssami, Effect of implementing digital breast tomosynthesis (DBT) instead of mammography on population screening outcomes including interval cancer rates: results of the Trento DBT pilot evaluation, Breast 50 (2020) 135–140.
[13] N. Houssami, D. Bernardi, F. Caumo, S. Brunelli, C. Fanto, M. Valentini, G. Romanucci, M.A. Gentilini, M. Zorzi, P. Macaskill, Interval breast cancers in the' screening with tomosynthesis or standard mammography' (STORM) population-based trial, Breast 38 (2018) 150–153.
[14] T. Hovda, A.S. Holen, K. Lang, J.L. Albertsen, H. Bjorndal, S.H.B. Brandal, K. K. Sahlberg, P. Skaane, P. Suhrke, S. Hofvind, Interval and consecutive round breast Cancer after digital breast tomosynthesis and synthetic 2D mammography versus standard 2D digital mammography in BreastScreen Norway, Radiology (2019), 191337.
[15] E.S. McDonald, A. Oustimov, S.P. Weinstein, M.B. Synnestvedt, M. Schnall, E. F. Conant, Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast Cancer screening, JAMA Oncol. 2 (6) (2016) 737–743.
[16] P. Skaane, S. Sebuødegård, A.I. Bandos, D. Gur, B.H. Østerås, R. Gullien, S. Hofvind, Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial, Breast Cancer Res. Treat. 169 (3) (2018) 489–496.
[17] S. Hofvind, A.S. Holen, H.S. Aase, N. Houssami, S. Sebuødegård, T.A. Moger, I. S. Haldorsen, L.A. Akslen, Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial, Lancet Oncol. 20 (6) (2019) 795–805.
[18] F. Caumo, M. Zorzi, S. Brunelli, G. Romanucci, R. Rella, L. Cugola, P. Bricolo, C. Fedato, S. Montemezzi, N. Houssami, Digital breast tomosynthesis with synthesized two-dimensional images versus full-field digital mammography for population screening: outcomes from the verona screening program, Radiology 287 (1) (2018) 37–46.
[19] T. Hovda, S.H.B. Brandal, S. Sebuødegård, A.S. Holen, H. Bjørndal, P. Skaane, S. Hofvind, Screening outcome for consecutive examinations with digital breast tomosynthesis versus standard digital mammography in a population-based screening program, Eur. Radiol. 29 (12) (2019) 6991–6999.
[20] L. Jiang, T. Ma, M.S. Moran, X. Kong, X. Li, B.G. Haffty, Q. Yang, Mammographic features are associated with clinicopathological characteristics in invasive breast cancer, Anticancer Res. 31 (6) (2011) 2327–2334.
[21] Y. Li, J. Cao, Y. Zhou, F. Mao, S. Shen, Q. Sun, Mammographic casting-type calcification is an independent prognostic factor in invasive breast cancer, Sci. Rep. 9 (1) (2019) 10544.
[22] L. Tabar, H.H. Tony Chen, M.F. Amy Yen, T. Tot, T.H. Tung, L.S. Chen, Y.H. Chiu, S.W. Duffy, R.A. Smith, Mammographic tumor features can predict long-term outcomes reliably in women with 1-14-mm invasive breast carcinoma, Cancer 101 (8) (2004) 1745–1759.
[23] H.S. Tsau, A.M. Yen, J.C. Fann, W.Y. Wu, C.P. Yu, S.L. Chen, S.Y. Chiu, L. Tabar, W. H. Kuo, H.H. Chen, K.J. Chang, Mammographic tumour appearance and triple-negative breast cancer associated with long-term prognosis of breast cancer death: a Swedish Cohort Study, Cancer Epidemiol. 39 (2) (2015) 200–208.
[24] F. Caumo, G. Romanucci, K. Hunter, M. Zorzi, S. Brunelli, P. Macaskill, N. Houssami, 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 Res. Treat. 170 (2) (2018) 391–397.
[25] P. Skaane, A.I. Bandos, R. Gullien, E.B. Eben, U. Ekseth, U. Haakenaasen, M. Izadi, I.N. Jebsen, G. Jahr, M. Krager, L.T. Niklason, S. Hofvind, D. Gur, Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program, Radiology 267 (1) (2013) 47–56.
[26] C. Vijapura, L. Yang, J. Xiong, L.L. Fajardo, Imaging features of nonmalignant and malignant architectural distortion detected by tomosynthesis, AJR Am. J. Roentgenol. 211 (6) (2018) 1397–1404.
[27] M.L. Zuley, A.I. Bandos, M.A. Ganott, J.H. Sumkin, A.E. Kelly, V.J. Catullo, G. Y. Rathfon, A.H. Lu, D. Gur, Digital breast tomosynthesis versus supplemental diagnostic mammographic views for evaluation of noncalcified breast lesions, Radiology 266 (1) (2013) 89–95.
[28] F.J. Gilbert, L. Tucker, K.C. Young, Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool, Clin. Radiol. 71 (2) (2016) 141–150.
[29] H.S. Aase, A.S. Holen, K. Pedersen, N. Houssami, I.S. Haldorsen, S. Sebuødegård, B. Hanestad, S. Hofvind, 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, Eur. Radiol. 29 (3) (2019) 1175–1186.
[30] S. Hofvind, K. Tsuruda, G. Mangerud, A.K. Ertzaas, et al., The Norwegian Breast Cancer Screening Program, 1996-2016: Celebrating 20 Years of Organised Mammographic screening, Cancer in Norway 2016 - Cancer incidence, Mortality, Survival and Prevalence in Norway, Cancer Registry of Norway, Oslo, 2017.
[31] N. Moshina, H.S. Aase, A.S. Danielsen, I.S. Haldorsen, C.I. Lee, S. Zackrisson, S. Hofvind, Comparing screening outcomes for digital breast tomosynthesis and digital mammography by automated breast density in a randomized controlled trial: results from the To-Be trial, Radiology 297 (3) (2020) 522–531.
[32] C.J. D'Orsi, E.A. Sicles, E.B. Mendelson, E.A. Morris, et al., ACR BI-RADS Atlas, Breast Imaging Reporting and Data System, American College of Radiology, Reston, VA, 2013.
[33] B. Naume, E. Schlichting, H.P. Eikesdal, A.I. Hagen, E.S. Blix, National Recommendations of Diagostics and Treatment of Breast Cancer, Norwegian Directorate of Health, Oslo, 2019.
[34] A. Goldhirsch, E.P. Winer, A.S. Coates, R.D. Gelber, M. Piccart-Gebhart, B. Thurlimann, H.J. Senn, m. Panel, Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013, Ann. Oncol. 24 (9) (2013) 2206–2223.
[35] K. Johnson, S. Zackrisson, A. Rosso, H. Sartor, L.H. Saal, I. Andersson, K. Lang, Tumor characteristics and molecular subtypes in breast Cancer Screening with digital breast tomosynthesis: the malmo breast tomosynthesis screening trial, Radiology 293 (2) (2019) 273–281.
[36] A.J. Evans, S.E. Pinder, J.J. James, I.O. Ellis, E. Cornford, Is mammographic spiculation an independent, good prognostic factor in screening-detected invasive breast cancer? AJR Am. J. Roentgenol. 187 (5) (2006) 1377–1380.
[37] S. Liu, X.D. Wu, W.J. Xu, Q. Lin, X.J. Liu, Y. Li, Is there a correlation between the presence of a spiculated mass on mammogram and luminal a subtype breast cancer? Korean J. Radiol. 17 (6) (2016) 846–852.
[38] G. Mariscotti, M. Durando, N. Houssami, M. Fasciano, A. Tagliafico, D. Bosco, C. Casella, C. Bogetti, L. Bergamasco, P. Fonio, G. Gandini, Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories, Breast Cancer Res. Treat. 166 (3) (2017) 765–773.
[39] M. Bahl, J.A. Baker, E.N. Kinsey, S.V. Ghate, Architectural distortion on mammography: correlation with pathologic outcomes and predictors of malignancy, AJR Am. J. Roentgenol. 205 (6) (2015) 1339–1345.
[40] P. Skaane, A.I. Bandos, L.T. Niklason, S. Sebuodegard, B.H. Osteras, R. Gullien, D. Gur, S. Hofvind, Digital mammography versus digital mammography plus tomosynthesis in breast Cancer screening: the oslo tomosynthesis screening trial, Radiology 291 (1) (2019) 23–30.

Graphic design: Communication Division, UiB / Print: Skipnes Kommunikasjon AS

uib.no

ISBN: 9788230843642 (print)
9788230869352 (PDF)