Radiology

Digital Mammography versus Breast Tomosynthesis: Impact of Breast Density on Diagnostic Performance in Population-based Screening

Bjørn Helge Østerås, MSc • Anne Catrine T. Martinsen, PhD • Randi Gullien, MSc • Per Skaane, MD, PhD

From the Department of Diagnostic Physics (B.H.Ø., A.C.T.M.) and Division of Radiology and Nuclear Medicine (R.G., P.S.), Oslo University Hospital, Building 20, Gaustad, PO Box 4959, Nydalen, 0424 Oslo, Norway; and Institute of Clinical Medicine (B.H.Ø., P.S.) and Department of Physics (A.C.T.M.), University of Oslo, Oslo, Norway. Received March 1, 2019; revision requested April 15; final revision received June 4; accepted June 14. **Address correspondence to** B.H.Ø. (e-mail: *bjorn.helge.osteras@ous-hf.no*).

P.S. received equipment and funding for additional case interpretations related to the Oslo Tomosynthesis Screening Trial from Hologic.

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

See also the editorial by Fuchsjäger and Adelsmayr in this issue.

Radiology 2019; 293:60–68 • https://doi.org/10.1148/radiol.2019190425 • Content code: BR

Background: Previous studies comparing digital breast tomosynthesis (DBT) to digital mammography (DM) have shown conflicting results regarding breast density and diagnostic performance.

Purpose: To compare true-positive and false-positive interpretations in DM versus DBT according to volumetric density, age, and mammographic findings.

Materials and Methods: From November 2010 to December 2012, 24 301 women aged 50–69 years (mean age, 59.1 years ± 5.7) were prospectively included in the Oslo Tomosynthesis Screening Trial. Participants received same-compression DM and DBT with independent double reading for both DM and DM plus DBT reading modes. Eight experienced radiologists rated the images by using a five-point scale for probability of malignancy. Participants were followed up for 2 years to assess for interval cancers. Breast density was assessed by using automatic volumetric software (scale, 1–4). Differences in true-positive rates, false-positive rates, and mammographic findings were assessed by using confidence intervals (Newcombe paired method) and P values (McNemar and χ^2 tests).

Results: The true-positive rate of DBT was higher than that of DM for density groups (range, 12%–24%; $P < .001$ for density scores of 2 and 3, and $P > .05$ for density scores of 1 and 4) and age groups (range, 15%–35%; $P < .05$ for all age groups), mainly due to the higher number of spiculated masses and architectural distortions found at DBT ($P < .001$ for density scores of 2 and 3; $P < .05$ for women aged 55–69 years). The false-positive rate was lower for DBT than for DM in all age groups (range, −0.6% to −1.2%; $P < .01$) and density groups (range, −0.7 to −1.0%; $P < .005$) owing to fewer asymmetric densities ($P \leq .001$), except for extremely dense breasts (0.1%, $P = .82$).

Conclusion: Digital breast tomosynthesis enabled the detection of more cancers in all density and age groups compared with digital mammography, especially cancers classified as spiculated masses and architectural distortions. The improvement in cancer detection rate showed a positive correlation with age. With use of digital breast tomosynthesis, false-positive findings were lower due to fewer asymmetric densities, except in extremely dense breasts.

Online supplemental material is available for this article.

The sensitivity of digital mammography (DM) is lower in women with dense breasts than in those with lower breast density (1). Breast density is also associated with higher false-positive rates and recall rates (2) due to superposition of normal glandular tissue that can mimic cancer. The woman's age has an impact on mammography screening as breast density decreases (3) and cancer incidence increases. The distribution of cancers shifts toward less-aggressive slower-growing cancers with increasing age (4). It has been shown that mammography screening has a lower sensitivity (1) and higher false-positive rate (2) among younger women.

Digital breast tomosynthesis (DBT) generates pseudo three-dimensional (3D) images where a single section of anatomy is in focus. The rest is blurred, with greater magnitude proportional to the distance from the focus plane. The screening performance of DBT for specific density and age groups may be different from that of DM, as DBT potentially can reduce masking and resolve superposition of breast tissue. Prospective (5–11) and retrospective (12–18) studies have shown that the integration of DBT improves the cancer detection or recall rates for both fatty and dense breasts and in age groups relevant for mammography screening. Data are limited in almost entirely fatty and extremely dense breasts. Two large studies compared DBT and DM in women with extremely dense breasts, with one study finding an increased cancer detection rate with DBT (5) and the other finding similar rates for DBT and DM (13). Therefore, there is a need for more data from large prospective trials.

Abbreviations

BI-RADS = Breast Imaging Reporting and Data System, CI = confidence interval, DBT = digital breast tomosynthesis, DM = digital mammography

Summary

For digital breast tomosynthesis compared with digital mammography, true-positive rates were higher and false-positive rates were lower for all volumetric breast density categories (except for extremely dense breasts) and age groups (ages 50–69 years).

Key Results

- The true-positive rate with digital breast tomosynthesis (DBT) was higher than that with digital mammography (DM) in all volumetric density groups (range, 12%–24%; $P < .001$ in women with scattered fibroglandular and heterogeneously dense breasts; $P > .05$ in women with almost entirely fatty and extremely dense breasts) and all age groups (range, 15%–35%; $P < .05$).
- The false-positive rate with DBT was lower than that with DM in all age groups (range, −0.6% to −1.2%; $P < .01$) and volumetric density groups (range, −0.7% to −1.0%; $P < .005$), except for women with extremely dense breasts (0.1%, $P = .82$).
- DBT showed a greater number of true-positive findings classified as spiculated masses or architectural distortions ($P < .001$ in women with scattered fibroglandular and heterogeneously dense breasts, $P < .05$ in women aged 55–69 years) and a reduction of false-positive findings classified as asymmetric densities ($P < .001$, except in women with extremely dense breasts).

Radiologists usually classify breasts into one of four Breast Imaging Reporting and Data System (BI-RADS) density categories (19). This method has considerable interobserver variability (20,21). Commercial software for automatic density classification has recently become available. Such software uses image processing and a physical model of the breast to calculate the woman's breast density objectively (22), thereby facilitating reproducible breast density stratification in the mammography screening.

The paired design of the prospective Oslo Tomosynthesis Screening Trial facilitates comparison of true- and false-positive interpretation between DM and DBT. Previous analysis showed an improvement in true- and false-positive rates with DBT (23). The benefit across density and age groups has not previously been analyzed in this cohort.

The aim of this study was to compare true- and false-positive interpretations in DM and DBT in prospective population-based screening according to volumetric density, age, and mammographic findings.

Materials and Methods

This prospective clinical trial (ClinicalTrials.gov NCT01248546) was approved by the regional ethics committee (reference number: 2010/144). Written informed consent was obtained from all participants. Hologic (Bedford, Mass) sponsored this study by providing equipment and financial support for additional readings. The authors had control of data and information submitted for publication. Five reports have been published on the Oslo Tomosynthesis Screening Trial, including two interim analyses comparing DM and DBT (n = 12 631) (24,25). After inclusion of all women, a study comparing two versions of synthetic DM (n = 24 901, some women were imaged twice) was published (26). After 2 years of follow-up to assess interval cancers, the sensitivity and specificity of DM plus DBT were compared with those from previous DM screening rounds (n = 24 301) (27) and for all screening arms (23). None of these reports have stratified results according to breast density or age, which is the goal of this preplanned analysis.

Study Cohort

From November 2010 to December 2012, 24 301 women (48 451 breasts) imaged at one breast center were included in the Oslo Tomosynthesis Screening Trial (Fig 1). The mean participant age ± standard deviation was 59.1 years ± 5.7. Women with pacemakers, women unable to stand, and women with breast implants were not included. The selection of women was solely based on the availability of radiographers and imaging systems. Recruitment was part of the population-based screening program BreastScreen Norway, which invites women aged 50–69 years to undergo biennial two-view screening. Eight experienced radiologists (including P.S.) with 2–31 years of experience in screening mammography (average, 16 years) participated in the trial. Before the trial, each radiologist received intensive personalized training with a set of 100 cancer-enriched cases. Details regarding study population and radiologist training are reported elsewhere (23,26).

Mammographic Imaging

Imaging was performed with three mammography systems (Selenia Dimensions, Hologic) by using the "combo" mode (single breast compression for DM and DBT). Craniocaudal and mediolateral oblique projections were acquired of both breasts. A standard screening setting (auto filter) was used, resulting in a mean average glandular radiation dose of 1.74 and 2.10 mGy for DM and DBT views, respectively (28).

Image Evaluation

Radiologists categorized their findings by using the BreastScreen Norway scale for the probability of cancer, as follows: 1, negative or definitely benign; 2, probably benign; 3, indeterminate; 4, probably malignant; and 5, malignant. A score of 2 or higher was classified as a positive mammographic finding, and these examinations were discussed at a consensus meeting (Fig 1). Four readers independently interpreted four study arms: two DM arms and two DM plus DBT arms. The workstation for each arm was in different rooms, and the patient's score for the respective arm was locked after closing the reading session. More details regarding study arms are reported elsewhere (23,27). Scores from DM arms were combined into a single score: two-dimensional (2D) double reading. If at least one DM arm had a positive score, 2D double reading was considered positive. Similarly, scores from DM plus DBT arms were combined into double reading 2D plus 3D. For positive scores, the radiologist classified the mammographic finding as mass (round, oval, irregular), spiculated mass, architectural distortion, asymmetric density, calcification, or calcification with density. All screening-detected cancers were classified at consensus.

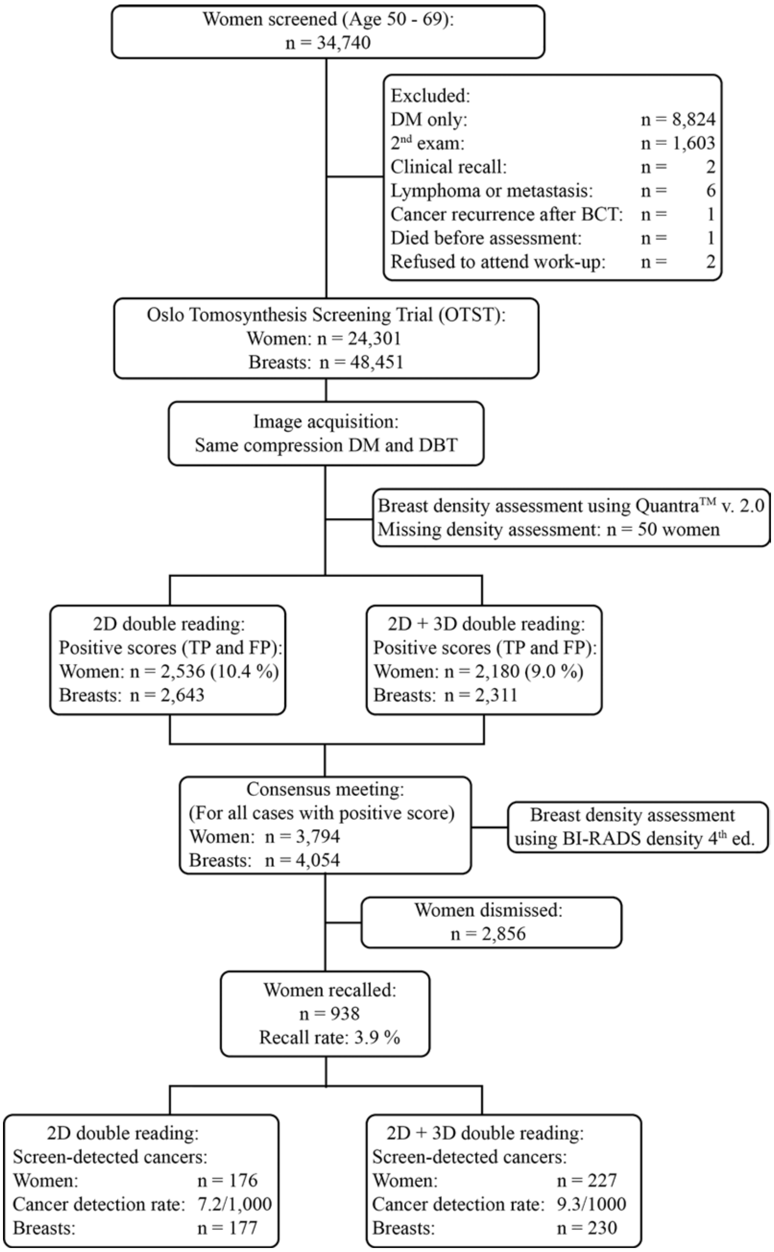

Figure 1: Flowchart of study population shows overall recall rate and cancer detection rate for independent two-dimensional (2D) and 2D plus three-dimensional (3D) double reading. BCT = breast-conserving therapy, BI-RADS = Breast Imaging Reporting and Data System, DBT = digital breast tomosynthesis, DM = digital mammography, FP = false positive, OTST = Oslo Tomosynthesis Screening Trial, TP = true positive, 2D = study arms using DM only, 2D+3D = study arms using DM plus DBT.

Breast Density Assessment

Volumetric breast density was calculated automatically by using commercial software (Quantra, version 2.0; Hologic). This information was not shown to readers (Fig 1). The software uses the raw DM image, physical model of the radiographic imaging chain, and attenuation in adipose and fibroglandular tissue to estimate volumetric breast density (ratio of fibroglandular to total breast volume) (22). Results for each view are aggregated into woman-based scores. Volumetric density was mapped to a quantized density score that was similar to BI-RADS 4th edition density scores, as follows: 1, almost entirely fatty; 2, scattered fibroglandular densities; 3, heterogeneously dense; and 4, extremely dense.

At the consensus meetings at least two or three of the eight participating radiologists assessed breast density in consensus according to BI-RADS 4th edition density (Fig 1) (19).

Statistical Analysis

Differences in distributions of breast density were assessed by using the χ^2 test (tabulate and chi2 commands, Stata, version 15.1; StataCorp, College Station, Tex). Agreement in density assessment between volumetric and BI-RADS density was assessed by using κ statistics with quadratic weights (kap, wgt [w2] commands; Stata) and Spearman correlation coefficients (Spearman command; Stata). The 95% confidence intervals (CIs) were estimated by using bootstrapping with 10 000 replacements (bootstrap command; Stata).

The 95% CIs in differences in proportions were estimated by using the Newcombe method for paired proportions (the Newcombe method for unpaired proportions was used as a conservative estimate where one modality found all screening-detected cancers).

The McNemar test (mcci command; Stata) was used when comparing differences in true- and false-positive rates for 2D and 2D plus 3D. For all analyses, $P < .05$ was considered indicative of a statistically significant difference.

We calculated the true-positive rate difference for 2D plus 3D and 2D with respect to age by using linear regression (regstat command; Matlab, Natick, Mass). Associated 95% CIs were calculated by using bootstrapping with 10 000 replacements (Matlab).

Differences in proportions of mammographic finding classifications in 2D and 2D plus 3D were evaluated by using the χ^2 test.

Results

Automatically Calculated Breast Density Distributions

With use of volumetric density, 65% of women (15 785 of 24 251) were considered to have nondense breasts (density 1 and 2) and 35% (8466 of 24 251) were considered to have dense breasts (density 3 and 4). The density distribution was different in all age groups ($P < .001$), shifting toward lower breast density with age (Table 1). The density distribution was different in women with positive scores ($P < .001$) and screening-detected cancers ($P = .002$) compared with all women. In addition, the density distribution was different between screening-detected and interval cancers when density was measured with use of volumetric (quantized) density ($P = .03$).

Table 1: Volumetric Breast Density according to Age, Positive Mammographic Score, and Screening-detected and Interval Cancers

Subgroup	No. of Women	No. of Women with Missing Density Measurements*	Volumetric Density Grade†			
			1	2	3	4
All included women	24 301	50	11.8 (2863/24 251)	53.3 (12 922/24 251)	27.4 (6645/24 251)	7.5 (1821/24 251)
Age 50–54 y	6508	10	6.2 (401/6498)	45.2 (2939/6498)	35.7 (2318/6498)	12.9 (840/6498)
Age 55–59 y	6693	14	11.2 (751/6679)	53.0 (3538/6679)	28.7 (1917/6679)	7.1 (473/6679)
Age 60–64 y	5578	10	15.0 (837/5568)	56.7 (3157/5568)	22.9 (1275/5568)	5.4 (299/5568)
Age 65–69 y	5522	16	15.9 (874/5506)	59.7 (3288/5506)	20.6 (1135/5506)	3.8 (209/5506)
Women with positive mammographic score	3794	4	7.8 (295/3790)	49.4 (1871/3790)	33.3 (1262/3790)	9.6 (362/3790)
Women with screening-detected cancers‡	230	1	7.0 (16/229)	46.7 (107/229)	36.2 (83/229)	10.0 (23/229)
Women with interval cancers§	51	0	0.0 (0/51)	33 (17/51)	49 (25/51)	18 (9/51)

* Density measurement was missing for one woman with bilateral screening-detected cancer.

† Data are percentages, with raw data in parentheses. Volumetric density was obtained with software (Quantra, Hologic).

‡ Four screening detected cancers was bilateral.

§ One interval cancer was bilateral.

Table 2: Correlation between BI-RADS and Volumetric Density

BI-RADS Density	Volumetric Density			
	1	2	3	4
I	115	189	6	2
II	177	1305	246	11
III	3	377	918	175
IV	0	0	92	174

Note.—Data are numbers of women. Breast Imaging Reporting and Data System (BI-RADS) density was obtained with BI-RADS 4th edition (19). Volumetric ("quantized") density was obtained with software (Quantra, Hologic). Interobserver agreement between BI-RADS density and volumetric density was substantial (κ = 0.69 [95% confidence interval: 0.67, 0.71]; Spearman correlation coefficient: 0.70 [95% confidence interval: 0.68, 0.72]).

Table 2 shows the Spearman correlation and interobserver agreement between volumetric density and BI-RADS density for women with positive mammographic scores (n = 3790). The interobserver agreement (κ value) was substantial with correlation ($P < .001$) among the two measures of breast density assessment.

True- and False-Positive Interpretations according to Density

Radiologists detected more cancers using 2D plus 3D compared with 2D double reading for all breast densities (Table 3). The breast-based true-positive rate was higher by 13% (two of 15; P = .50) for almost entirely fatty breasts, 33% (26 of 79; $P < .001$) for breasts with scattered fibroglandular densities, 30% (19 of 64; $P < .001$) for heterogeneously dense breasts, and 28% (five of 15; P = .06) for extremely dense breasts. The 95% CIs for the difference in true-positive rate between 2D plus 3D and 2D overlap for all density categories.

Table E1 (online) shows true-positive interpretations stratified according to BI-RADS density, with similar results as stratification with volumetric ("quantized") density.

Radiologists reported fewer false-positive scores using 2D plus 3D compared with 2D double reading for women with all densities, except those with extremely dense breasts. The breast-based false-positive rate was lower by 23% (45 of 197; P = .004) for almost entirely fatty breasts, 21% (252 of 1224; $P < .001$) for breasts with scattered fibroglandular densities, and 12% (94 of 815; P = .004) for heterogeneously dense breasts. The breast-based false-positive rate was higher by 2% (five of 229; P = .82) for extremely dense breasts.

True- and False-Positive Interpretations according to Age

The number of true-positive scores was higher for all age strata for 2D plus 3D compared with 2D double reading (Table 3). The improvement in the true-positive rate was 18% (eight of 44 breasts; P = .008) for ages 50–54 years, 19% (nine of 48 breasts; P = .02) for ages 55–59 years, 33% (15 of 46 breasts; $P < .001$) for ages 60–64 years, and 54% (21 of 39 breasts; $P <$ .001) for ages 65–69 years. The 95% CIs for the difference in the true-positive rate overlap for all age strata. Still, linear regression

Table 3: Breast-based True- and False-Positive Interpretations for 2D and 2D Plus 3D Double Reading according to Volumetric Density and Age

	No. of Breasts		True-Positive Interpretations				False-Positive Interpretations			
Parameter	With SDC*	Without SDC†	No. with 2D	No. with 2D plus 3D	Difference‡	*P* Value	No. with 2D	No. with 2D plus 3D	Difference‡	*P* Value
All women	234	48 217	177	230	53 (22.7) [17.0, 28.6]	<.001	2466	2081	−385 (−0.80) [−1.03, −0.57]	<.001
Volumetric density§										
1 and 2	125	31 334	94	122	28 (22.4) [14.3, 30.8]	<.001	1421	1124	−297 (−0.95) [−1.21, −0.69]	<.001
3 and 4	107	16 787	82	106	24 (22.4) [14.1, 31.4]	<.001	1044	955	−89 (−0.53) [−0.96, −0.10]	.02
1	17	5681	15	17	2 (11.8) [−8.5, 34.3]‖	.50	197	152	−45 (−0.79) [−1.33, −0.26]	.004
2	108	25 653	79	105	26 (24.1) [15.1, 33.3]	<.001	1224	972	−252 (−0.98) [−1.28, −0.69]	<.001
3	84	13 182	64	83	19 (22.6) [12.9, 32.9]	<.001	815	721	−94 (−0.71) [−1.19, −0.24]	.004
4	23	3605	18	23	5 (21.7) [3.0, 41.9]‖	.06	229	234	5 (0.14) [−0.83, 1.11]	.82
Age										
50–54 y	52	12 952	44	52	8 (15.4) [5.3, 27.5]‖	.008	901	800	−101 (−0.78) [−1.27, −0.29]	.002
55–59 y	59	13 296	48	57	9 (15.3) [3.4, 27.5]	.02	619	542	−77 (−0.58) [−1.01, −0.15]	.009
60–64 y	63	11 049	46	61	15 (23.8) [11.7, 36.1]	<.001	521	388	−133 (−1.20) [−1.66, −0.76]	<.001
65–69 y	60	10 920	39	60	21 (35.0) [22.6, 47.6]‖	<.001	425	351	−74 (−0.68) [−1.12, −0.24]	.003

Note.—There were 2643 true- and false-positive interpretations with two-dimensional (2D) double reading and 2311 with 2D and three-dimensional (3D) double reading. SCD = screening-detected cancer.

* One woman with bilateral cancer had missing density. Therefore, analysis stratified according to density is missing for two breasts.

† Density measurements were missing in 49 women (96 breasts). Therefore, analysis stratified according to density is missing for 96 breasts.

‡ Data are numbers of interpretations, with percentages in parentheses and 95% confidence intervals in brackets.

§ Volumetric (“quantized”) density was obtained with software (Quantra, Hologic).

‖ Determined with the Newcombes method, unpaired.

showed an improvement of 6.7% (95% CI: 1.8%, 11.6%; $P < .01$) for every 5 years, with an r^2 of 0.87 (95% CI: 0.14, 0.99).

The number of false-positive interpretations was lower for all age strata with use of 2D plus 3D compared with 2D double reading (Table 3). The false-positive rate was lower by 11% (101 of 901 breasts; $P = .001$) for ages 50–54 years, 12% (77 of 619 breasts; $P = .009$) for ages 55–59 years, 26% (133 of 521 breasts; $P < .001$) for ages 60–64 years, and 17% (74 of 425 breasts; $P = .003$) for ages 65–69 years.

Age and Density Adjustments

Tables E2–E4 (online) show the age-adjusted difference in true- and false-positive interpretations stratified according to volumetric (quantized) and BI-RADS density and the volumetric density–adjusted difference in true- and false-positive interpretations stratified according to age, respectively. The tables show only minor differences in estimates when adjusting for age or volumetric density.

True- and False-Positive Mammographic Features according to Density

The number of breast-based true-positive interpretations for 2D plus 3D and 2D double reading stratified according to mammographic finding and volumetric (quantized) density is shown in Figure 2. Most additional cancers found in the 2D plus 3D analysis were classified as a spiculated mass or architectural distortion (Figs 3, E1 [online]). For these lesions, differences were 25% in almost entirely fatty breasts (two of eight [$P = .50$], with 10 detected with 2D plus 3D and eight detected with 2D), 43% in breasts with scattered fibroglandular densities (20 of 46 [$P < .001$], with 66 detected with 2D plus 3D and 46 detected with 2D), 40% in heterogeneously dense

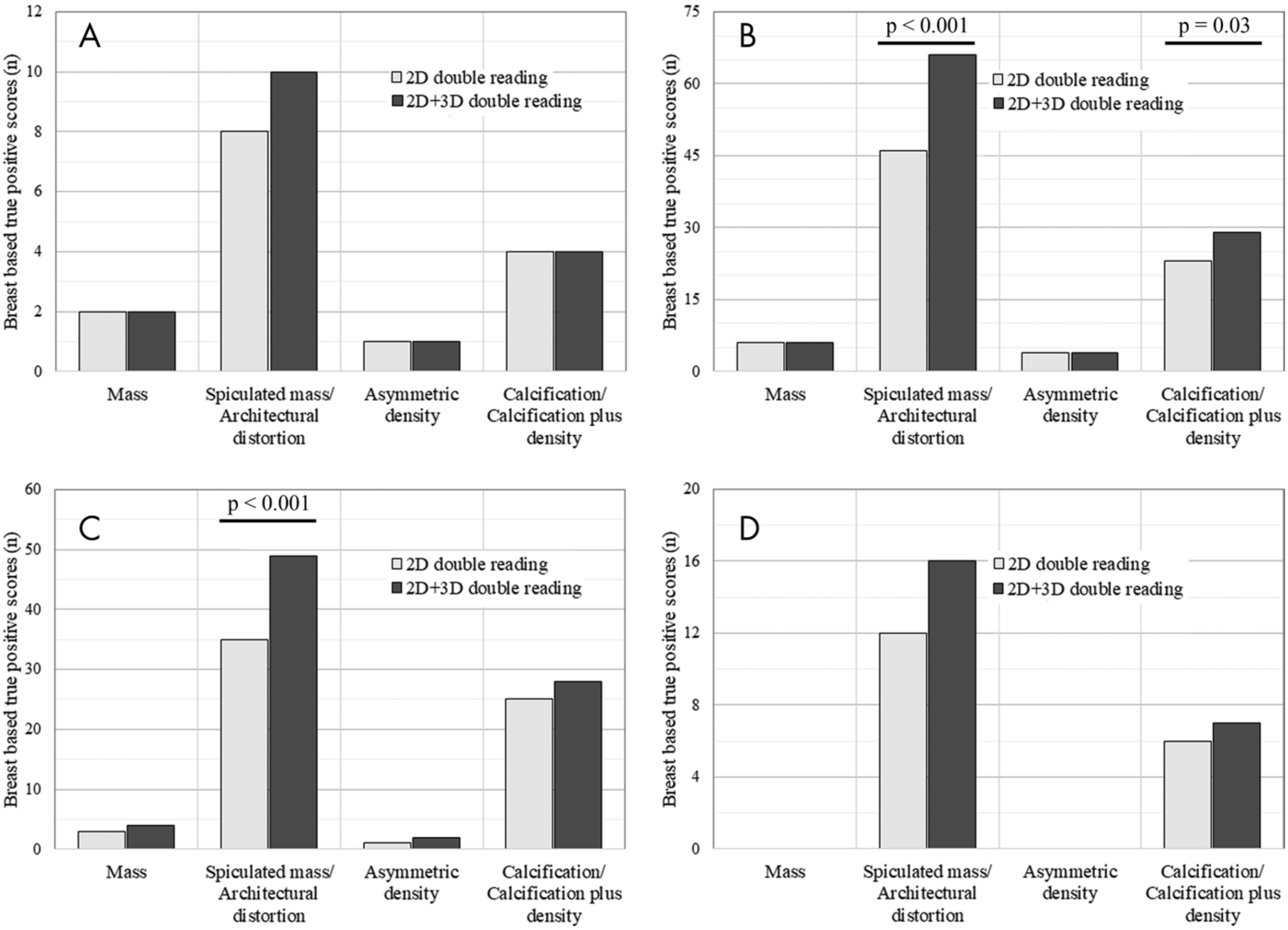

Figure 2: Bar charts show breast-based true-positive findings for two-dimensional (2D) and 2D plus three-dimensional (3D) double reading stratified according to volumetric ("quantized") density in, *A*, entirely fatty breasts, *B*, scattered fibroglandular breasts, *C*, heterogeneously dense breasts, and, *D*, extremely dense breasts. *P* values are given for significant differences. One woman with bilateral cancer had missing density values. Quantized density was obtained with software (Quantra, Hologic).

breasts (14 of 35 [$P < .001$], with 49 detected with 2D plus 3D and 35 detected with 2D), and 33% in extremely dense breasts (four of 12 [$P = .13$], with 16 detected with 2D plus 3D and 12 detected with 2D). In addition, more cancers classified as calcification and/or calcification plus density were found in women with scattered fibroglandular breasts (26%, six of 23 [$P = .03$], with 29 detected with 2D plus 3D and 23 detected with 2D).

The number of breast-based false-positive interpretations for 2D plus 3D and 2D double reading stratified according to the mammographic finding reported by the readers and volumetric (quantized) density is shown in Figure E2 (online). There was a reduction in false-positive interpretations classified as asymmetric densities with use of 2D plus 3D for women in all density categories (Fig E3 [online]), except for women with extremely dense breasts. The reduction was 51% in almost entirely fatty breasts (32 of 63 [$P = .001$], with 31 false-positive interpretations with 2D plus 3D and 63 with 2D), 58% in breasts with scattered fibroglandular densities (231 of 397 [$P < .001$], with 166 false-positive interpretations with 2D plus 3D and 397 with 2D), 52% in heterogeneously dense breasts (111 of 215 [$P < .001$], with 104 false-positive interpretations with 2D plus 3D and 215 with 2D), and 27% in extremely dense breasts (13 of 48 [$P = .15$], with 35 false-positive interpretations with 2D plus 3D and 48 with 2D). There was a lower amount of false-positive findings classified as spiculated mass and/or architectural distortions and a higher amount of false-positive findings classified as calcifications and/or calcification plus density for 2D plus 3D compared to 2D.

True- and False-Positive Mammographic Features according to Age

The number of breast-based true-positive interpretations for 2D plus 3D and 2D double reading stratified according to mammographic findings and age is shown in Figure E4 (online). The number of true-positive findings was higher in 2D plus 3D only for spiculated masses and/or architectural distortions for all age groups. The magnitude of the increase was 18% for age 50–54 years (five of 28 [$P = .06$], with 33 true-positive interpretations for 2D plus 3D and 28 with 2D), 21% for age 55–59 years (seven of 33 [$P = .04$], with 40 true-positive interpretations for 2D plus 3D and 33 for 2D), 60% for age 60–64 years (12 of 20 [$P = .002$], with 32 true-positive interpretations for 2D plus 3D and 20 for 2D), and 76% for age 65–69 years

(16 of 21 [$P < .001$], with 37 true-positive interpretations for 2D plus 3D and 21 with 2D). The number of breast-based false-positive interpretations for 2D plus 3D and 2D stratified according to age and mammographic finding reported by the readers is shown in Figure 4. There was a similar difference in type of false-positive findings across all age categories.

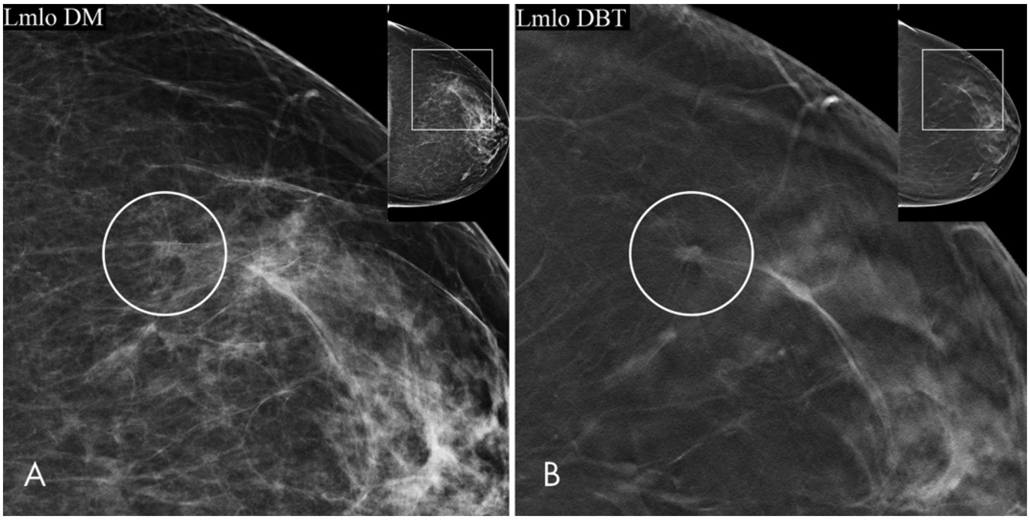

Figure 3: Screening mammograms (mediolateral oblique [Lmlo] views) obtained with, *A*, digital mammography (DM) and, *B*, digital breast tomosynthesis (DBT) in left breast of 68-year-old woman with fatty breast (volumetric density score of 2 with software [Quantra, Hologic] and Breast Imaging Reporting and Data System density category II) show a spiculated mass (in circle). Positive scores were given only by the DBT readers; the cancer was overlooked by both two-dimensional mammography readers. Histologic examination revealed an 8-mm tubular carcinoma. As the textures masking the tumor in the DM image are small, out-of-plane blurring effectively renders the masking textures invisible on the DBT images, leaving the in-plane tumor and spiculations visible. In addition, a dark halo image artifact in the tube movement direction helps highlight the tumor on the DBT image (29).

Discussion

We compared cancer detection rates and false-positive findings with digital breast tomosynthesis (DBT) and digital mammography (DM) stratified according to density and age, as previous studies have shown conflicting results. The results of our study show that adding DBT to DM in screening yields more cancers in women of all density categories (range, 12%–24%; $P < .001$ in women with scattered fibroglandular and heterogeneously dense breasts, $P > .05$ in women with almost entirely fatty and extremely dense breasts) and age groups (range, 15%–35%; $P < .05$). Most additional cancers manifest as spiculated masses or architectural distortions ($P < .001$ in women with scattered fibroglandular and heterogeneously dense breasts; $P < .05$ in women aged 55–69 years). Improvement in cancer detection with DBT showed a positive correlation ($P < .01$) with age (50–69 years). The false-positive rate was lower for women in most density categories (range, −0.7% to −1.0%; $P < .005$) and age groups (range, −0.6% to −1.2%, $P < .01$). The false-positive rate was not lower in women with extremely dense breasts (0.1%, $P = .82$), mostly due to lower number of asymmetric densities ($P < .001$).

Studies have shown improvement in cancer detection by using DBT over DM in fatty and dense breasts (5,9,11–13,15), in agreement with our results. The Malmö trial (5) indicated improvement using DBT for women with all densities, whereas another retrospective study (13) indicated similar performance, except for women with extremely dense breasts. Our results agree with those from the Malmö trial (5), which found improved cancer detection for women with extremely dense breasts and most additional cancers manifesting as spiculated masses or architectural distortions. Image texture from normal tissue (eg, fibroglandular tissue) can mask tumor spiculations and mass at DM (29). DBT removes out-of-plane fine structures, leaving in-plane tumor and spiculations visible. It has been shown that DBT requires at least some amount of peritumoral fat to be effective in dense breasts (30). In addition, the anatomic noise of structures larger than 2 mm is almost identical in DBT and DM images (31). This indicates that DBT does not allow radiologists to "see though" dense breast parenchyma but removes fine textures masking tumor spiculations in women of all breast densities. Our study showed that radiologists detected more spiculated masses with DBT compared with DM as the woman's age increases. An explanation is as age increases, the proportion of less-aggressive slower-growing cancers increases (4). Such small low-grade tumors tend to manifest as spiculated masses (6), which are better visualized with DBT. Other studies have shown increased cancer detection for all age groups in women aged 50–69 years (5,8,9,11,14,15); to our knowledge, no studies have shown positive correlation with age.

Studies have shown a reduction in the recall rate using DBT for women in all breast density categories (9,12,13,15,16,18). Our results and those from another Norwegian trial (10) differ from the results of these studies, with no difference in false-positive or recall rates for women with extremely dense breasts. However, we found a reduction in false-positive findings in heterogeneously dense breasts; the other study did not (10). The differences in the false-positive or recall rate in women with extremely dense breasts might be explained by the large difference in recall rate between countries. Other studies (32,33) have shown recall reduction mainly due to asymmetric densities, similar to our results. Superimposition of glandular tissue creates pseudo-lesions in DM, which DBT often resolves as glandular tissue is depicted in different sections. Unlike the Malmö trial (7), we detected fewer false-positive findings classified as spiculated mass or architectural distortions. We found more false-positive findings with calcifications, which might be due to highlighting of calcifications by using synthetic 2D imaging.

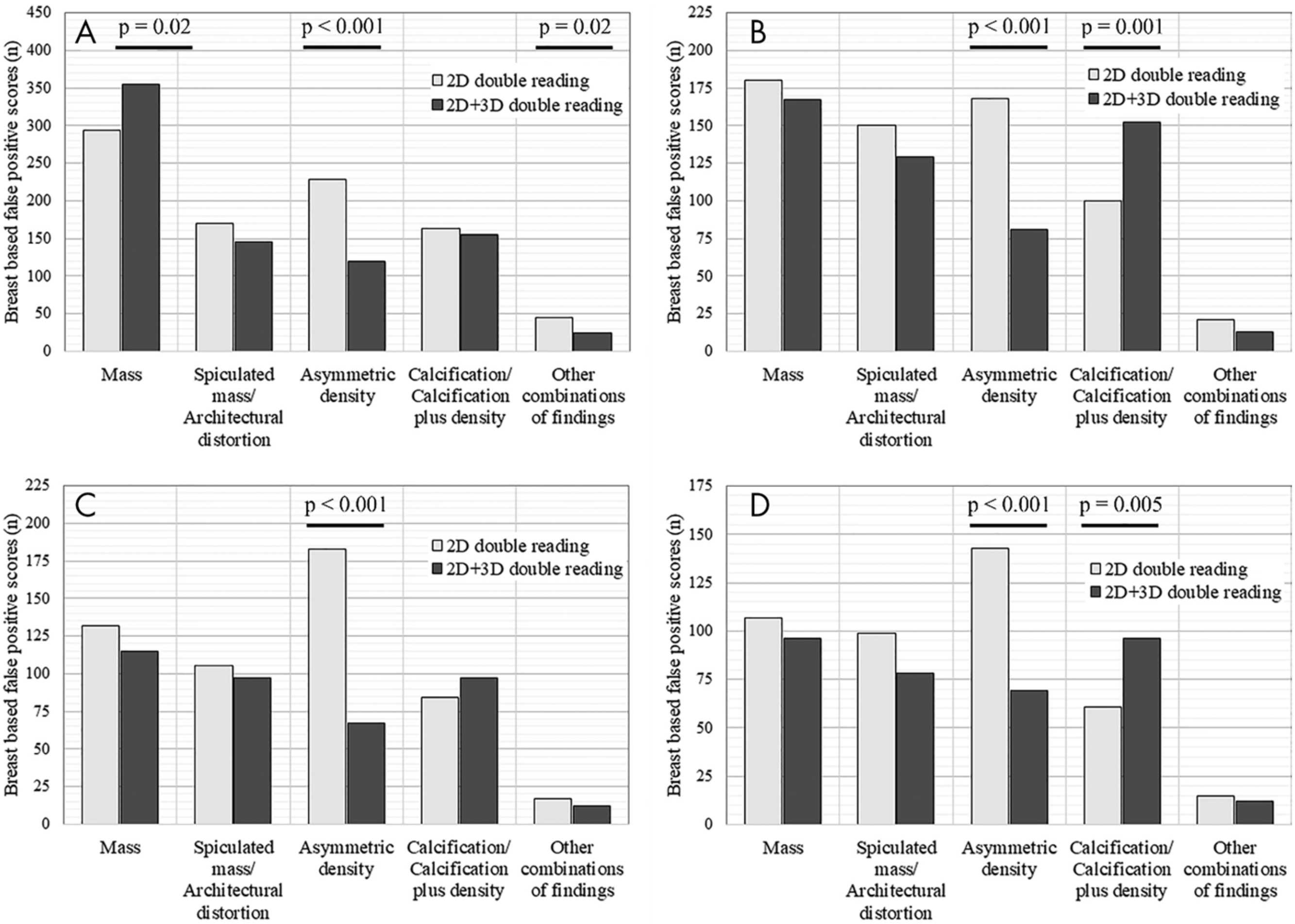

Figure 4: Bar charts show breast-based false-positive findings for two-dimensional (2D) and 2D plus three-dimensional (3D) double reading stratified according to patient age: *A*, 50–54 years, *B*, 55–59 years, *C*, 60–64 years, and, *D*, 65–69 years. *P* values are given for significant differences.

Density was assessed with use of volumetric and BI-RADS density. BI-RADS density has large interobserver variability, making it potentially advantageous to use volumetric density (20,21). Our study and a previous analysis (34) showed that agreement of BI-RADS and volumetric density is limited in almost entirely fatty and extremely dense breasts. Most discrepant classifications will be borderline cases (34), resulting in similar comparison of true- and false-positive rates by using either density measure. If volumetric density was used, two-thirds of the women would be classified as having fatty breasts, compared with half if using BI-RADS density (19,34).

Our study has limitations. It was a single-institution trial. We used the 4th edition of BI-RADS for breast density categorization because this was the standard scale when the Oslo Tomosynthesis Screening Trial started. In addition, Quantra version 2.0 maps volumetric density into categories according to the BI-RADS 4th edition density. Newer versions of Quantra use the BI-RADS 5th edition as reference. In the 5th edition, breasts are classified into a higher category if an area is dense and can obscure lesions. The 5th edition might be more associated with a reduction in sensitivity in both DM and DBT as very dense areas can obscure cancers.

In conclusion, mammography screening using digital breast tomosynthesis (DBT) depicted more cancers in all density and age groups compared with digital mammography (DM) owing to the higher number of cancers classified as spiculated masses and architectural distortions at DBT. Improvement in cancer detection showed a positive correlation with age. The number of false-positive findings with DBT was lower than that with DM due to fewer asymmetric densities, except in extremely dense breasts.

Author contributions: Guarantor of integrity of entire study, B.H.Ø.; 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, B.H.Ø., P.S.; clinical studies, B.H.Ø., R.G., P.S.; statistical analysis, B.H.Ø., A.C.T.M.; and manuscript editing, all authors.

Disclosures of Conflicts of Interest: **B.H.O.** disclosed no relevant relationships. **A.C.T.M.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: Oslo University Hospital has a research collaboration with GE Heathcare on CT. Other relationships: disclosed no relevant relationships. **R.G.** disclosed no relevant relationships. **P.S.** Activities related to the present article: institution received equipment and funding for additional case interpretations from Hologic. Activities not related to the present article: received payment for lectures including service on speakers bureaus from Hologic. Other relationships: disclosed no relevant relationships.

References

1. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138(3):168–175.
2. Lehman CD, White E, Peacock S, Drucker MJ, Urban N. Effect of age and breast density on screening mammograms with false-positive findings. AJR Am J Roentgenol 1999;173(6):1651–1655.
3. Sprague BL, Gangnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst 2014;106(10):1–6.
4. Howlader N, Altekruse SF, Li CI, et al. US incidence of breast cancer subtypes defined by joint hormone receptor and HER2 status. J Natl Cancer Inst 2014;106(5):1–8.
5. 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.
6. Lång K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: results from the Malmö Breast Tomosynthesis Screening Trial, a population-based study. Eur Radiol 2016;26(1):184–190.
7. Lång K, Nergården M, Andersson I, Rosso A, Zackrisson S. False positives in breast cancer screening with one-view breast tomosynthesis: an analysis of findings leading to recall, work-up and biopsy rates in the Malmö Breast Tomosynthesis Screening Trial. Eur Radiol 2016;26(11):3899–3907.
8. Bernardi D, Macaskill P, Pellegrini M, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. Lancet Oncol 2016;17(8):1105–1113.
9. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 2013;14(7):583–589.
10. Aase HS, Holen ÅS, 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.
11. 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.
12. 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.
13. 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.
14. Rafferty EA, Rose SL, Miller DP, et al. Effect of age on breast cancer screening using tomosynthesis in combination with digital mammography. Breast Cancer Res Treat 2017;164(3):659–666.
15. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. Breast Cancer Res Treat 2016;156(1):109–116.
16. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology 2013;269(3):694–700.
17. Sharpe RE Jr, Venkataraman S, Phillips J, et al. Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast tomosynthesis into a population-based screening program. Radiology 2016;278(3):698–706.
18. Alsheik NH, Dabbous F, Pohlman SK, et al. Comparison of resource utilization and clinical outcomes following screening with digital breast tomosynthesis versus digital mammography: findings from a learning health system. Acad Radiol 2019;26(5):597–605.
19. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS): mammography. 4th ed. Reston, Va: American College of Radiology, 2003.
20. Sprague BL, Conant EF, Onega T, et al. Variation in mammographic breast density assessments among radiologists in clinical practice: a multicenter observational study. Ann Intern Med 2016;165(7):457–464.
21. Østerås BH, Martinsen ACT, Brandal SHB, et al. Classification of fatty and dense breast parenchyma: comparison of automatic volumetric density measurement and radiologists' classification and their inter-observer variation. Acta Radiol 2016;57(10):1178–1185.
22. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. Volumetric breast density estimation from full-field digital mammograms. IEEE Trans Med Imaging 2006; 25(3):273–282.
23. Skaane P, Bandos AI, Niklason LT, et al. Digital mammography versus digital mammography plus tomosynthesis in breast cancer screening: the Oslo Tomosynthesis Screening Trial. Radiology 2019;291(1):23–30.
24. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 2013;267(1):47–56.
25. Skaane P, Bandos AI, Gullien R, et al. 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 2013;23(8):2061–2071.
26. Skaane P, Bandos AI, Eben EB, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. Radiology 2014;271(3):655–663.
27. Skaane P, Sebuødegård S, Bandos AI, et al. Performance of breast cancer screening using digital breast tomosynthesis: results from the prospective population-based Oslo Tomosynthesis Screening Trial. Breast Cancer Res Treat 2018;169(3):489–496.
28. 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.
29. Nakashima K, Uematsu T, Itoh T, et al. Comparison of visibility of circumscribed masses on Digital Breast Tomosynthesis (DBT) and 2D mammography: are circumscribed masses better visualized and assured of being benign on DBT? Eur Radiol 2017; 27(2):570–577.
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. 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.
32. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. Radiology 2015;274(1):85–92.
33. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. Radiology 2015;274(2):337–342.
34. Østerås BH, Martinsen ACT, Brandal SHB, et al. BI-RADS Density Classification From Areometric and Volumetric Automatic Breast Density Measurements. Acad Radiol 2016;23(4):468–478.