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*).

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.

© RSNA, 2020

Online supplemental material is available for this article.

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 mammramphy, 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 28 749 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 of no special type.

We included data on weight and height from a questionnaire used in BreastScreen Norway from 2006 to 2016 (22). Breast

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/m²)*	25 (23–28)	25 (23–28)
Without information	4378	4499
Breast volume (cm³)*	844 (576–1171)	848 (571–1190)
Without information	121	86
Fibroglandular volume (cm³)*	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.

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 χ² 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/andersskyrud/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 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
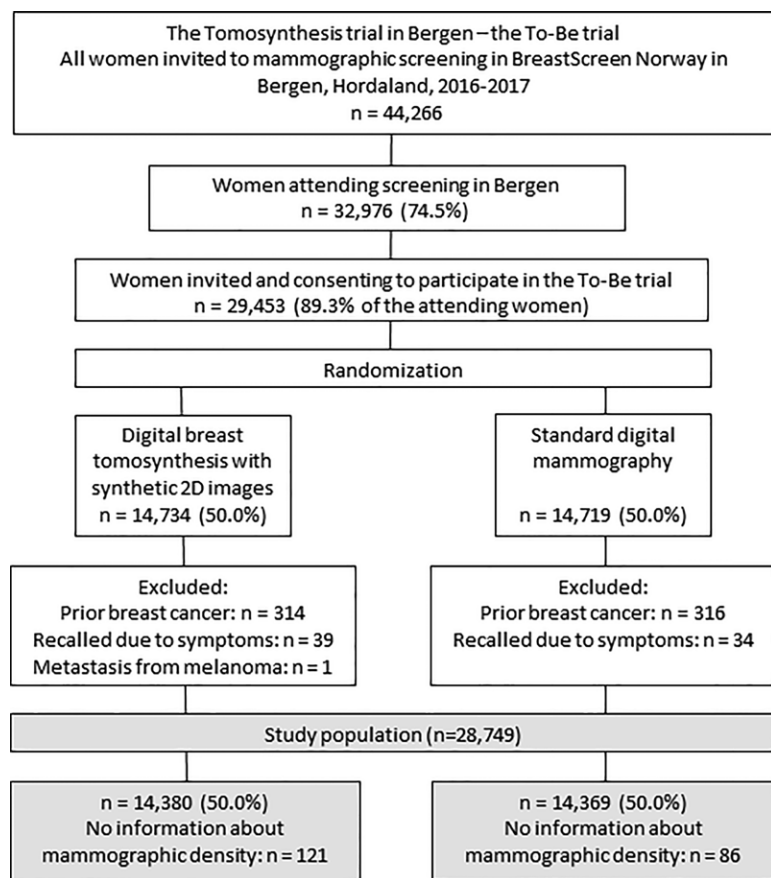
Figure 1: Flowchart shows exclusion criteria and the final study sample for women screened with digital breast tomosynthesis including synthesized two-dimensional (2D) mammography and standard digital mammography.

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/m² (IQR, 23–28 kg/m²), 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 cm³ (IQR, 576–1171 cm³) for DBT+SM and 848 cm³ (IQR, 571–1190 cm³) for DM. In the DBT+SM arm, median fibroglandular volume was 39.8 cm³ (IQR, 29.6–55.0 cm³), whereas it was 42.9 cm³ (IQR, 32.0–58.5 cm³) 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 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).
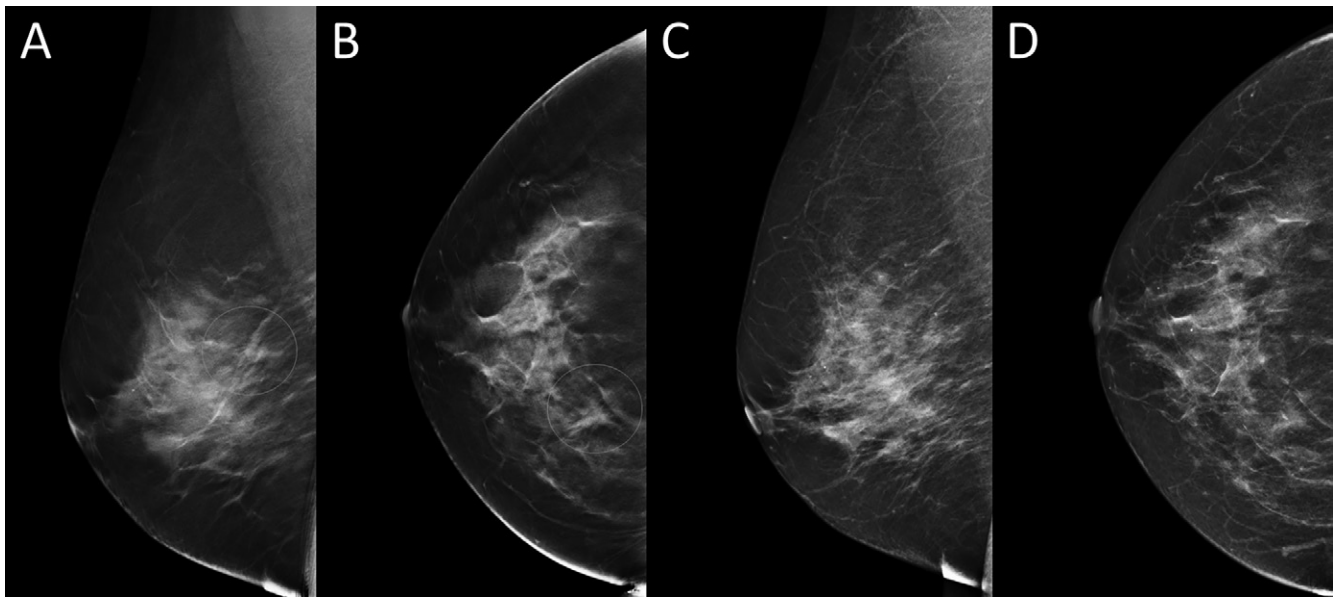
Figure 2: *A,* Right mediolateral oblique and, *B,* craniocaudal digital breast tomosynthesis images at 1-mm plane, and, *C,* right mediolateral oblique and, *D,* craniocaudal synthetic two-dimensional images in a 59-year-old woman with high breast density (Breast Imaging Reporting and Data System 3). 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.
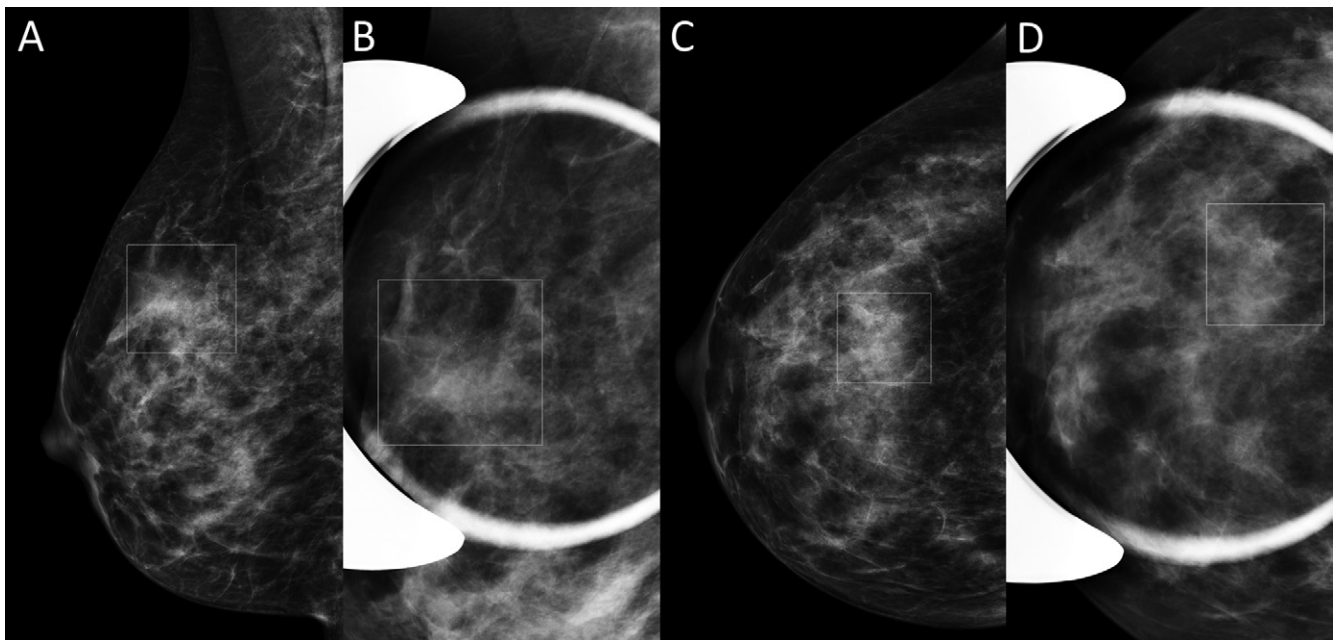
Figure 3: *A,* Right mediolateral oblique and, *C,* craniocaudal digital mammography images with, *B, D,* spot magnification in a 56-year-old woman with high breast density (Breast Imaging Reporting and Data System 3). Both readers detected calcifications in the central upper part of the right breast (□). Histologic examination revealed a 25-mm invasive ductal carcinoma of no special type (grade 3, luminal B).

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 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).

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.

Sensitivity analyses using VBD quintiles verified increasing RRs of SDC in the DBT+SM models with increasing 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:
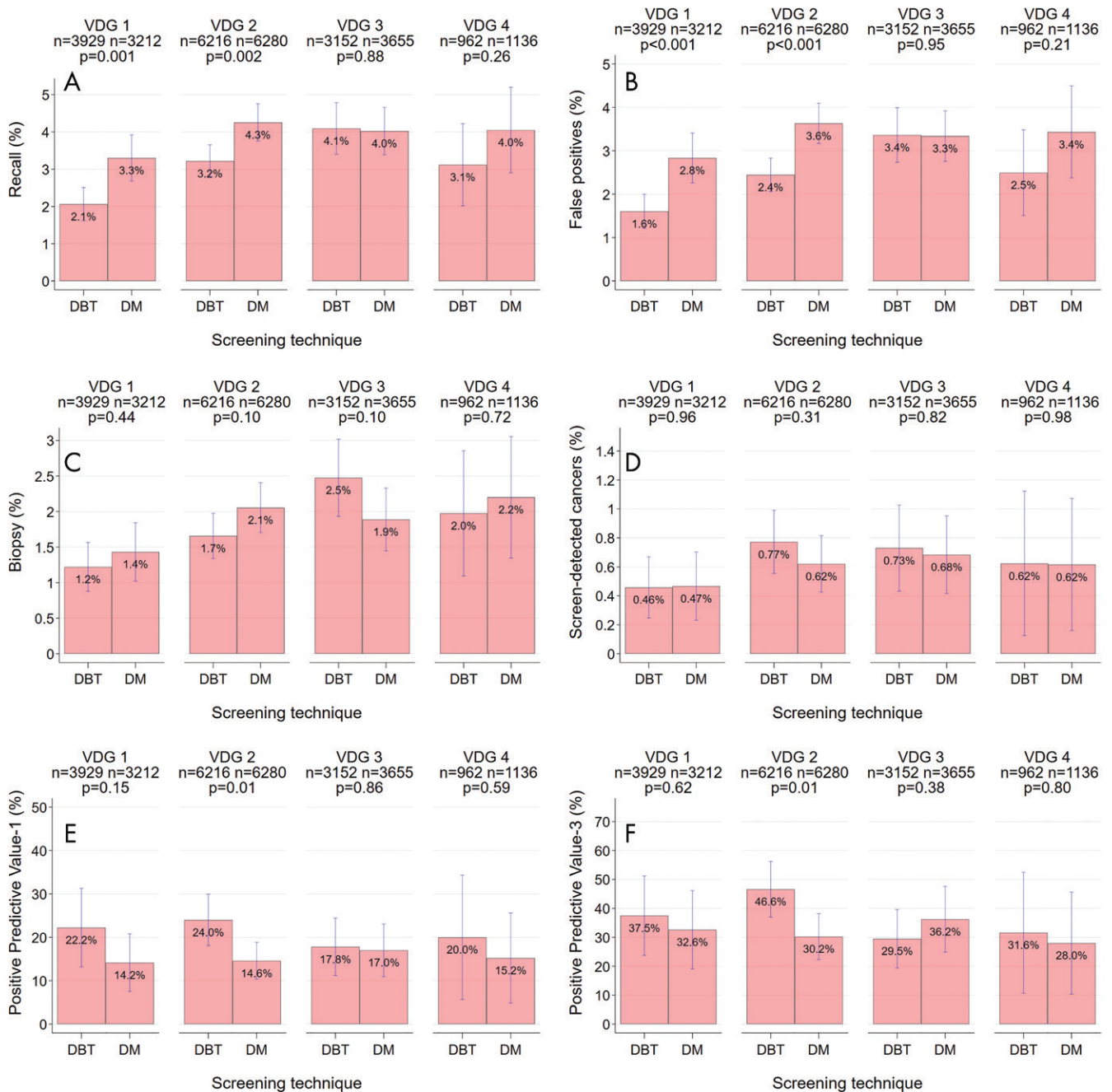
Figure 4: Rates with 95% confidence intervals of, *A*, recall, *B*, false-positive screening examinations, *C*, biopsy, *D*, screen-detected breast cancer, *E*, positive predictive value of recalls, and, *F*, positive predictive value of biopsy for digital breast tomosynthesis (DBT) that includes synthesized 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.

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 screen-

ing examinations, biopsies, and screen-detected breast cancer, and histopathologic findings for DBT+SM versus DM by automated density. Women with nondense breasts had a 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;

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

Characteristic	Relative Risk of Recall for DBT+SM						Relative Risk 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	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 (cm³)	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

Characteristic	Relative Risk of FP for DBT+SM						Relative Risk of FP for DM					
	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 (cm³)	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.

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 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

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

Characteristic	RR of SDC for DBT+SM						RR of SDC for DM					
	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.

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.

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 (MB-TST): 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.