Automated Volumetric Analysis of Mammographic Density in a Screening Setting: Worse Outcomes for Women with Dense Breasts

Nataliia Moshina, MD, PhD • *Sofie Sebuødegård, MSc* • *Christoph I. Lee, MS, MD* • *Lars A. Akslen, MD, PhD* • *Kaitlyn M. Tsuruda, MSc* • *Joann G. Elmore, MD, MPH* • *Solveig Hofvind, PhD*

From the Cancer Registry of Norway, Oslo, Norway (N.M., S.S., K.M.T., S.H.); Departments of Radiology (C.I.L.) and Medicine (J.G.E.), University of Washington School of Medicine, Seattle, Wash; Department of Clinical Medicine, Section for Pathology, Centre for Cancer Biomarkers (CCBIO), Bergen, Norway (L.A.A.); Department of Pathology, Haukeland University Hospital, Bergen, Norway (L.A.A.); and Oslo Metropolitan University, Faculty of Health Science, Oslo, Norway (S.H.). Received December 20, 2017; revision requested February 2, 2018; final revision received March 26; accepted April 12. **Address correspondence to** S.H. (e-mail: *Solveig.Hofvind@kreftregisteret.no*).

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

See also the editorial by Philpotts in this issue.

Purpose: To describe screening outcomes from BreastScreen Norway stratified by volumetric breast density (VBD).

Materials and Methods: This retrospective study included data from 107 949 women aged 50–69 years (mean age ± standard deviation, 58.7 years ± 5.6) who underwent 307 015 screening examinations from 2007 to 2015. Automated software classified mammographic density as nondense (VBD <7.5%) or dense (VBD ≥7.5%). Rates and distributions of screening outcomes (recall, biopsy, screen-detected and interval breast cancer, positive predictive values of recall and of needle biopsy, sensitivity, specificity, and histopathologic tumor characteristics) were analyzed and stratified by density. Tests of proportions, including propensity score and *t* tests, were used.

Results: In 28% (87 021 of 307 015) of the screening examinations, the breasts were classified as dense. Recall rates for women with nondense versus dense breasts were 2.7% (5882 of 219 994) and 3.6% (3101 of 87 021); biopsy rates were 1.1% (2359 of 219 994) and 1.4% (1209 of 87 021); rates of screen-detected cancer were 5.5 (1210 of 219 994) and 6.7 (581 of 87 021) per 1000 examinations; and rates of interval breast cancer were 1.2 (199 of 165 324) and 2.8 (185 of 66 674) per 1000 examinations, respectively (*P* < .001 for all). Sensitivity was 82% (884 of 1083) for nondense breasts and 71% (449 of 634) for dense breasts, whereas specificity was 98% (160 973 of 164 440) and 97% (64 250 of 66 225), respectively (*P* < .001 for both). For screen-detected cancers, mean tumor diameter was 15.1 mm and 16.6 mm (*P* = .01), and lymph node–positive disease was found in 18% (170 of 936) and 24% (98 of 417) (*P* = .02) of women with nondense and dense breasts, respectively.

Conclusion: Screening examinations of women with dense breasts classified by using automated software resulted in higher recall rate, lower sensitivity, larger tumor diameter, and more lymph node–positive disease compared with women with nondense breasts.

©RSNA, 2018

Online supplemental material is available for this article.

Stratified breast cancer screening among women with average risk is now a recommended practice in the United States, with some suggesting tailored regimens based on breast density (1,2). Women with mammographically dense breasts have a higher risk of breast cancer and a higher risk of missed cancers than do those with nondense breasts (3–5). The sensitivity of mammography is reported to be as low as 60% for women with extremely dense breasts (5–7). The higher risk of breast cancer among women with dense breasts is etiologically related to biologic mechanisms (8), whereas the superimposition of breast tissue leads to a masking effect whereby breast cancers can go undetected (9). Breast cancers are detected in dense breasts at a later stage and have less favorable tumor characteristics than do those detected in nondense breasts (7,10,11). Population-based mammographic screening may therefore be less effective for women with dense breasts than for those with nondense breasts.

The majority of studies about mammographic density have focused on the association between subjective density assessments and breast cancer risk (3–5). Only a few studies have used objective density measures for risk estimation, and even fewer studies have evaluated quantitative breast density measurement with regard to screening performance (7,11,12). The most commonly used method for classification of mammographic density is the radiologist's subjective interpretation using the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS) (13). However, no reference standard exists for breast density determination. To eliminate inter- and intrareader variability regarding density categorization, automated software has been developed to provide quantitative density assessments in real time (12,14).

In some places, supplemental screening with US and MRI and/or more frequent screening are options for women with dense breasts (4,15,16). In the United States, more than half of the states have enacted breast density legislation (17). In some places, this legislation mandates that women should be informed about their breast density or that additional

Abbreviations

BI-RADS = Breast Imaging Reporting and Data System, HER2 = human epidermal growth factor receptor 2, VBD = volumetric breast density, VDG = Volpara density grade

Summary

By using automated volumetric mammographic density assessment software, the authors identified higher recall and biopsy rates and higher odds of screen-detected and interval breast cancer for screening examinations of women with volumetrically dense versus nondense breasts.

Implications for Patient Care

- Automated volumetric breast density measurements may be considered a future standard for breast cancer screening, ensuring an objective density classification.

- By using automated volumetric breast density assessment software, screening examinations of women with dense breasts are associated with higher rates of recall and biopsy, and higher odds of screening-detected and interval breast cancers than are examinations of women with nondense breasts.

- By using automated volumetric breast density software, less than one in three women in a population-based screening program will be classified as having dense breasts.

imaging could be beneficial. However, supplemental screening for women with dense breasts is not currently recommended by any major societies or organizations (1,18).

Our study objective was to describe population-based screening performance and outcomes stratified by volumetric mammographic density. Specifically, we aimed to determine the rates of recall and biopsy, rates of cancer detection, positive predictive values, sensitivity, specificity, histopathologic tumor characteristics, odds of breast cancer, and predicted numbers of breast cancer cases based on volumetric breast density (VBD) categories. Our findings will help inform how automated volumetric density categorization will change population-based screening performance and outcomes under a more objective paradigm for breast density measurement. We hypothesized that screening examinations of women with high VBD were associated with less favorable screening outcomes compared with those of women with low breast density.

Materials and Methods

Our study was approved by the data protection official for research at Oslo University Hospital (Oslo, Norway) and the regional committee for medical and health research ethics. Neither the authors nor the study received any funding or support from industry.

Data for this retrospective study were extracted from databases at the Cancer Registry of Norway and deidentified prior to analyses. It has been mandatory to report information about neoplasms to the Cancer Registry of Norway since 1952 (19). Breast cancer completeness has been estimated to be close to 100%, and 99% of the cases are morphologically verified. According to the Cancer Registry Regulations, information about screening examinations performed in BreastScreen Norway—the national screening program for breast cancer—can be used for quality assurance and

research if the women have not actively opted out. About 2% of women attending the program have refused such data usage (20). The Cancer Registry of Norway coordinates BreastScreen Norway, which started in 1996 and offers biennial mammography to approximately 600 000 women aged 50–69 years (20). A unique 11-digit personal identification number assigned to all residents of Norway ensures 100% completeness of the invitations to the program. Attendance rate for each round of screening is approximately 75% and about 85% of all the invited women have attended the program at least once.

The program provides independent double reading performed by two breast radiologists; the readers are totally blinded to the interpretation scores of one another. Each breast is assigned a score of 1–5 by each radiologist to indicate mammographic findings (1, negative for malignancy; 2, probably benign; 3, intermediate suspicion of malignancy; 4, probably malignant; 5, high suspicion of malignancy). If either radiologist assigns a score of 2 or higher, then a consensus or arbitration meeting is held to determine whether to call the woman back for further assessment (recall). The recall rate in the program is about 3% (21). All radiologists are required to undergo training to start and continue screen reading (21). The experiences in screen reading for the radiologists included in this study varied from first-year faculty to those with more than 20 years of experience.

Study Sample

Data from women screened in Rogaland and Hordaland counties as a part of BreastScreen Norway during the study period of January 1, 2007 to December 31, 2015 ($n = 109\,821$ women and $329\,179$ screening examinations) constituted the study population (Fig 1).

We excluded information about 20 672 women and 22 164 screening examinations (6.8% [22 164 of 329 179] of the screening examinations) without information about VBD. Our final study population consisted of 107 949 women aged 50–69 years, with mean age at time of screening of 58.3 years $\pm$ 5.7 (standard deviation). The women underwent an average of 2.8 screening examinations during the study period, accounting for 307 015 examinations in total. Results of sensitivity analyses on screening outcome for women with and without density data did not differ significantly, except for positive predictive value of biopsies (Table E1 [online]).

Analyses pertaining to interval breast cancers, sensitivity, and specificity included information from women who underwent screening and had follow-up data for 2 years (96 052 women who underwent 231 998 examinations) to ensure sufficient follow-up time for detection of interval cancer (24 months).

Equipment Used and Assessment of VBD

All examinations were performed with full-field digital mammography (Senographe DS or Senographe Essential; GE Medical Systems SCS, Buc, France). Continuous measures of fibroglandular volume (absolute density), breast volume, and VBD (percent density) were obtained from raw data by using an automated software (Volpara, version 1.5.1; Volpara Solutions, Wellington, New Zealand) (22). The density values represented the

average value for a screening examination, which typically consisted of four images (craniocaudal and mediolateral oblique views of each breast).

This software used the maximum VBD value per examination to classify density into a four-category scale, or Volpara density grade (VDG). Examinations with VBD less than or equal to 4.5% were classified as VDG1, those with 4.5%–7.49% as VDG2, those with 7.5%–15.49% as VDG3, and those with greater than or equal to 15.5% as VDG4 (Figure 2). These volumetric density categories are considered analogous to the BI-RADS density categories (13). The correlation between radiologists' BI-RADS density scores and automated density assessment was previously reported to be moderate (12). We present results based on a binary classification of VDG as follows: VDG1 or VDG2 (VBD <7.5%) versus VDG3 or VDG4 (VBD ≥7.5%), hereafter referred to as nondense and dense, respectively. Results for the four VDG categories are provided in Figure E1 (online).

Figure 1: Flowchart of patient population in the study.

Variables of Interest

Screening history was based on women's personal screening history, recorded by the Cancer Registry of Norway. Prevalent examination was defined as the women's first examination in BreastScreen Norway, regular subsequent examination as those less than 752 days since last examination, and irregular subsequent examination as those greater than or equal to 752 days since last examination.

Individual breast cancer risk factors were captured by using a questionnaire sent with screening invitations (23). Information from 62% (66 817 of 107 949) of the women in the study population was available. Weight, height, use of hormonal therapy, number of pregnancies lasting greater than 6 months, and first- or second-degree family history of breast cancer were risk factors included in our analysis. Body mass index was calculated as weight divided by squared height and we defined less than 20.0 kg/m² as underweight, 20.0–24.9 kg/m² as healthy weight, 25.0–29.9 kg/m² as overweight, and greater than or equal to 30 kg/m² as obese. Use of hormonal therapy was categorized as never user, past user, current user of estrogen alone (eg, estradiol and estriol), current user of combined estrogen-progestin (estradiol and norethisterone acetate), or unspecified (current user of unspecified hormonal therapy or incomplete information on past use). The number of pregnancies lasting greater than 6 months was categorized as none, one to two, or at least three, and a positive family history was defined as a mother, sister, daughter, or grandmother diagnosed with breast cancer at any age.

Recall rate was defined as the percentage of screening examinations resulting in a call back for further assessment because of abnormal mammographic findings, whereas the biopsy rate was related to screening examinations resulting in a needle biopsy. A screen-detected breast cancer was defined as breast cancer (ductal carcinoma in situ or invasive breast cancer) diagnosed after a recall. An interval breast cancer was defined as a breast cancer diagnosed
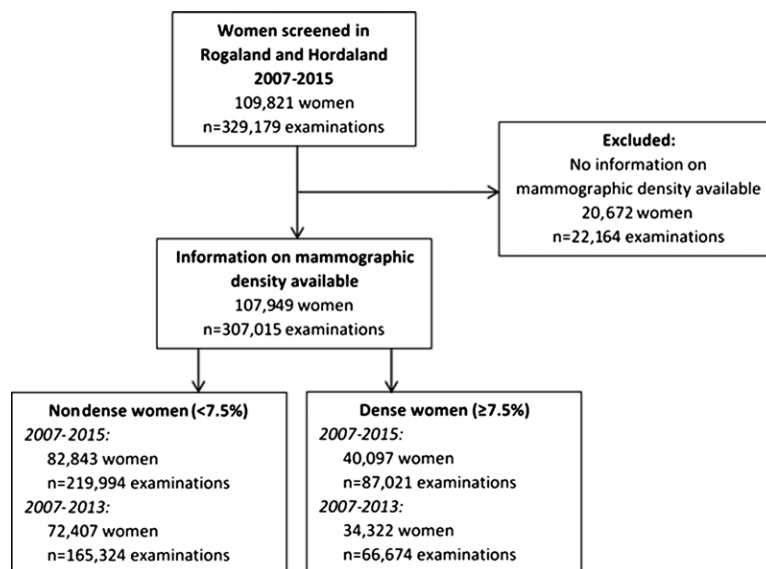
within 24 months of a negative screening examination or within 6–24 months of a false-positive screening result. Detection rates for screen-detected and interval breast cancer were presented per 1000 screening examinations. Positive predictive value was defined as the percentage of screen-detected breast cancers among recalls (hereafter, PPV-1) and as the percentage of screen-detected breast cancers among recalls leading to needle biopsy (hereafter, PPV-3). Sensitivity was defined as the percentage of screen-detected breast cancers among screen-detected and interval breast cancers. Specificity was defined as the percentage of true-negative screening examinations among false-positive and true-negative examinations.

Data about tumor characteristics included histopathologic type, tumor diameter (in millimeters), histologic grade, and lymph node involvement, as well as hormonal receptor (estrogen, estrogen receptor [ER], and progesterone, progesterone receptor [PR]) and human epidermal growth factor receptor 2 (HER2) status, which were used to determine immunohistochemical subtype classifications: luminal A–like (ER positive, PR positive, and HER2 negative), luminal B–like HER2 negative (ER positive, PR negative, and HER2 negative), luminal B–like HER2 positive (ER positive, PR positive, and HER2 positive), HER2 positive (ER negative, PR positive, and HER2 positive), and triple negative (ER negative, PR negative, and HER2 negative) (24).

Statistical Analyses

Our analysis was performed at the examination level. Distributions of the variables of interest were stratified by automated volumetric density (nondense vs dense breasts) and mode of detection (screen-detected or interval breast cancer). We tested for statistically significant differences between groups by using *t* tests and tests of proportions. We also performed a propensity score analysis to account for differences between the dense and nondense group.

The propensity score weights were estimated by fitting a logistic regression model on density, adjusted for age at screening, screening location, screening history, number of pregnancies, body mass index, use of hormonal therapy, number of pregnancies

Table 1: Recall and Biopsy Rates, Rates of Screen-detected and Interval Breast Cancers, and Positive Predictive Values of Recalls Due to Abnormal Mammographic Findings and Needle Biopsies among Women Who Participated in BreastScreen Norway by VBD

Parameter	Nondense (VBD <7.5%)	Dense (VBD ≥7.5%)	P Value	All Examinations (VBD, 0.1%–52%)
Screen-detected breast cancer				
No. of examinations[*]	219 994	87 021	…	307 015
Recalled	5882	3101	…	8983
Recall rate	2.7 (2.6, 2.7)	3.6 (3.4, 3.7)	<.0001	2.9 (2.9, 3.0)
Biopsies	2359	1209	…	3568
Biopsy rate	1.1 (1.0, 1.1)	1.4 (1.3, 1.5)	<.0001	1.2 (1.1, 1.2)
DCIS	237	152	…	389
Invasive	973	429	…	1402
DCIS + invasive	1210	581	…	1791
Rate of screen-detected breast cancer per 1000 screening examinations				
DCIS	1.1 (0.9, 1.2)	1.8 (1.5, 2.1)	<.0001	1.3 (1.2, 1.4)
Invasive	4.4 (4.2, 4.7)	4.9 (4.5, 5.4)	.06	4.6 (4.3, 4.8)
DCIS + invasive	5.5 (5.2, 5.8)	6.7 (6.1, 7.2)	.0001	5.8 (5.6, 6.1)
Recalls due to abnormal mammographic findings (PPV-1)				
DCIS	4.0 (3.5, 4.5)	4.9 (4.1, 5.7)	.053	4.3 (3.9, 4.8)
Invasive	16.5 (15.6, 17.5)	13.8 (12.6, 15.1)	.0008	15.6 (14.9, 16.4)
DCIS + invasive	20.6 (19.5, 21.6)	18.7 (17.4, 20.1)	.04	19.9 (19.1, 20.8)
Recalls resulting in needle biopsies (PPV-3)				
DCIS	10.1 (8.8, 11.3)	12.6 (10.7, 14.4)	.02	10.9 (9.9, 12.0)
Invasive	41.2 (39.3, 43.2)	35.5 (32.8, 38.2)	.0009	39.3 (37.7, 40.9)
DCIS + invasive	51.3 (49.3, 53.3)	48.1 (45.2, 50.9)	.07	50.2 (48.6, 51.8)
Interval breast cancer				
No. of examinations	165 324	66 674	…	231 998
DCIS	11	7	…	18
Invasive	188	178	…	366
DCIS + invasive	199	185	…	384
Rate of interval breast cancer per 1000 screening examinations[*]				
DCIS	0.07 (0.03, 0.11)	0.11 (0.03, 0.18)	.34	0.08 (0.05, 0.12)
Invasive	1.1 (1.0, 1.3)	2.7 (2.3, 3.1)	<.0001	1.6 (1.4, 1.8)
DCIS + invasive	1.2 (1.0, 1.4)	2.8 (2.4, 3.2)	<.0001	1.7 (1.5, 1.8)
Sensitivity[*]	81.6 (79.3, 83.9)	70.8 (67.3, 74.4)	<.0001	77.6 (75.6, 79.6)
Specificity[*]	97.9 (97.8, 98.0)	97.0 (96.9–97.1)	<.0001	97.6 (97.6, 97.7)

Note.—Unless otherwise specified, data are percentages, with 95% confidence intervals in parentheses. DCIS = ductal carcinoma in situ, PPV = positive predictive value, VBD = volumetric breast density.

[*] Information on interval breast cancer, sensitivity, and specificity was provided for women screened with full-field digital mammography in the program from 2007–2013 and were followed for 2 years.

lasting greater than 6 months, and first- or second-degree family history of breast cancer. Women with missing information on questionnaire data were included with mean values of the missing estimates, identified with a dummy variable. The inverse probability weight for individual i were defined as follows:

$$w_i = \frac{Z_i}{e_i} + \frac{(1 - Z_i)}{1 - e_i},$$

where e_i denote the propensity score for the i-th individual, and Z_i is an indicator variable for individual i having mammographically dense breasts.

The odds ratios of screen-detected and interval breast cancers associated with volumetric density were estimated by using generalized estimating equations with a logit link function and robust standard errors to account for within-woman correlation. Models were adjusted for age at screening, screening location, and screening history.

The number of cases of screen-detected and interval breast cancer per 1000 screening examinations were predicted and graphed by using locally weighted scatterplot smoothing to show the association between density and breast cancer, stratified by 5-year age groups. These predicted numbers were derived from generalized estimating equations models with

Table 2: Characteristics of Screen-detected Breast Cancers in BreastScreen Norway by VBD

Parameter	Nondense (VBD <7.5%)	Dense (VBD ≥7.5%)	P Value	All (VBD 0.1%–52%)
Screen-detected breast cancer	1210	581	…	1791
Histopathologic type				
Invasive ductal carcinoma	801 (66.2)	352 (60.6)	.02	1153 (64.4)
Invasive lobular carcinoma	88 (7.3)	49 (8.4)	.39	137 (7.7)
Other invasive	84 (6.9)	28 (4.8)	.08	112 (6.3)
Ductal carcinoma in situ	237 (19.6)	152 (26.2)	.002	389 (21.7)
Invasive breast cancer	973	429	…	1402
Tumor diameter (mm)				
Mean*	15.1 ± 9.3	16.6 ± 10.3	.009	15.6 ± 9.6
Median	13	14	…	13
<10	265 (28.0)	98 (23.6)	.08	363 (26.7)
10–19	460 (48.7)	193 (46.4)	.44	653 (48.0)
20–29	141 (14.9)	77 (18.5)	.10	218 (16.0)
>29	79 (8.4)	48 (11.5)	.06	127 (9.3)
Information not available	28	13	…	41
Histologic grade				
I	301 (32.1)	151 (36.0)	.15	452 (33.3)
II	421 (44.8)	183 (43.7)	.69	604 (44.5)
III	217 (23.1)	85 (20.3)	.25	302 (22.2)
Information not available	34	10	…	44
Lymph node status				
Positive	170 (18.2)	98 (23.5)	.02	268 (19.8)
Information not available	37	12	…	49
Subtypes				
Luminal A–like	615 (67.4)	250 (63.0)	.12	865 (66.0)
Luminal B–like HER2 negative	138 (15.1)	68 (17.1)	.36	206 (15.7)
Luminal B–like HER2 positive	59 (6.5)	43 (10.8)	.007	102 (7.8)
HER2 positive	39 (4.3)	11 (2.8)	.19	50 (3.8)
Triple negative	62 (6.8)	25 (6.3)	.74	87 (6.6)
Information not available	60	32	…	92

Note.—Data in parentheses are percentages. VBD = volumetric breast density, HER2 = human epidermal growth factor receptor 2.
* Data are means ± standard deviations.

density represented by cubic splines with six knots. Covariates in these models were the same as described above.

Because of multiple testing, we performed the Bonferroni correction, such that the adjusted P value < .0008 was considered to indicate statistical significance. Analyses were performed by using Stata (version 14.2; Stata, College Station, Tex). Data analyses and interpretation were performed by N.M. (with 4 years of expertise), S.S. (with 7 years of expertise), K.M.T. (with 6 years of expertise), and S.H. (with 20 years of expertise).

Results

Among the 307 015 screening examinations performed during 2007–2015, 14% were prevalent, 82% were regular subsequent, and 4% were irregular subsequent examinations. VBD measurements ranged from 1% to 52%; 72% (219 994 of 307 015) of examinations were classified as nondense and 28% (87 021 of 307 015) as dense (Table 1, Fig 1).

The recall rate was 2.7% (5882 of 219 994) for screening examinations of women with nondense breasts and 3.6% (3101 of 87 021) for those with dense breasts (P < .0001) (Table 1).

The rate of needle biopsy resulting from abnormal screening was 1.1% (2359 of 219 994) for screening examinations of those with nondense and 1.4% (1209 of 87 021) for those with dense breasts (P < .0001). A lower rate of breast cancer was observed for nondense versus dense breasts: 5.5 (1210 of 219 994) versus 6.7 per 1000 examinations (581 of 87 021) for screen-detected breast cancer, and 1.2 (199 of 165 324) versus 2.8 (185 of 66 674) per 1000 examinations for interval breast cancer, respectively (P < .001 for both). Sensitivity was 82% (884 of 1083) versus 71% (449 of 634) and specificity was 98% (160 973 of 164 440) versus 97% (64 250 of 66 225) for screening examinations of women with nondense and dense breasts, respectively (P < .0001 for both). Results of crude and propensity-weighted analyses did not differ for nondense versus dense breasts, except for the percentage of screen-detected breast cancers among recalls (PPV-1) and for the percentage of screen-detected breast cancers among recalls leading to needle biopsy (PPV-3) (Table E2 [online]).

Among women with screen-detected breast cancer, a lower proportion of ductal carcinoma in situ was found for those with nondense (20%, 237 of 1210) versus those with dense breasts

Table 3: Characteristics of Interval Breast Cancers Detected in BreastScreen Norway by VBD

Parameter	Nondense (VBD <7.5%)	Dense (VBD ≥7.5%)	P Value	All (VBD 0.1%–52%)
Interval breast cancer	199	185	…	384
Histopathologic type				
Invasive ductal carcinoma	133 (66.8)	150 (81.1)	.002	283 (73.7)
Invasive lobular carcinoma	30 (15.1)	19 (10.3)	.16	49 (12.8)
Other invasive	25 (12.6)	9 (4.9)	.008	34 (8.9)
Ductal carcinoma in situ	11 (5.5)	7 (3.8)	.42	18 (4.7)
Invasive breast cancer	188	178	…	366
Tumor diameter (mm)				
Mean*	25.3 ± 13.7	24.1 ± 14.4	.46	24.7 ± 14.1
Median	23	20	…	22
<10	16 (9.5)	13 (7.9)	.62	29 (8.7)
10–19	46 (27.2)	63 (38.4)	.03	109 (32.7)
20–29	56 (33.1)	46 (28.1)	.31	102 (30.6)
>29	51 (30.2)	42 (25.5)	.35	93 (27.9)
No information available	19	14	…	33
Histologic grade				
I	45 (14.1)	25 (14.8)	.86	50 (14.5)
II	77 (43.5)	70 (41.4)	.70	147 (42.5)
III	75 (42.4)	64 (43.8)	.79	149 (43.1)
No information available	11	9	…	20
Lymph node status				
Positive	72 (40.5)	72 (43.6)	.55	144 (42.0)
No information available	10	13	…	23
Subtypes				
Luminal A–like	86 (49.7)	88 (51.5)	.75	174 (50.6)
Luminal B–like HER2 negative	33 (19.1)	21 (12.3)	.08	54 (15.7)
Luminal B–like HER2 positive	18 (10.4)	19 (11.1)	.83	37 (10.8)
HER2 positive	7 (4.1)	17 (9.9)	.03	24 (7.0)
Triple negative	29 (16.8)	26 (15.2)	.69	55 (16.0)
Information not available	15	7	…	22

Note.—Data in parentheses are percentages. VBD = volumetric breast density, HER2 = human epidermal growth factor receptor 2.
* Data are means ± standard deviations.

(26%, 152 of 581; P = .002) (Table 2). A smaller mean tumor diameter and proportion of lymph node–positive tumors was found among those with nondense versus dense breasts (tumor diameter, 15.1 mm vs 16.6 mm; P = .009 and lymph node involvement, 18% [170 of 936] vs 24% [98 of 417]; P = .023). The proportion of luminal B–like HER2-positive tumors was 7% (59 of 913) for women with nondense and 11% (43 of 397) for women with dense breasts (P = .007).

Histopathologic tumor characteristics of interval breast cancer did not differ significantly for women with nondense versus dense breasts with regard to mean tumor diameter (25.3 mm vs 24.1 mm; P = .46), lymph node involvement (41% [72 of 178] vs 44% [72 of 165]; P = .55), and triple negative status (17% [29 of 173] vs 15% [26 of 171]; P = .69) (Table 3). Results of propensity-weighted analyses are shown in Tables E3 and E4 (online).

Additional results pertaining to screening performance measures and tumor characteristics for screen-detected and interval breast cancers stratified by four-group volumetric density categorization are available in Appendix E1 (online). Notably, screening examinations of women with fatty breasts (or VDG1) had a statistically significantly lower recall rate (2.2%, 2541 of 116 037), biopsy rate (0.9%, 1048 of 116 037), rate of screen-detected cancer (0.46%, 529 of 116 037), and rate of interval breast cancer (0.07%, 60 of 84 991) compared with those with the densest breasts (or VDG4) (recall rate, 3.6% [512 of 14 225]; biopsy rate, 1.6% [228 of 14 225]; rate of screen-detected cancer, 0.68% [96 of 14 225]; and rate of interval breast cancer, 0.31% [34 of 11 045]) (Table E5 [online]). The proportion of women classified as VDG4 was 5% (11 045 of 231 998) and tumors detected among these women contributed to 5.4% (96 of 1791) of the total number of the screen-detected cancers and 8.9% (34 of 384) of the interval breast cancers. Tumor characteristics were less favorable for women with VDG4 than for those classified as VDG1 (Tables E6 and E7 [online]).

Age, body mass index, use of hormonal therapy, number of pregnancies, and family history were significantly associated with volumetric density categories (Table 4). Notably, the average body mass index was higher for screening examinations of women with breasts classified as nondense compared with those classified as dense (27 kg/m² vs 23 kg/m²; P < .0001).
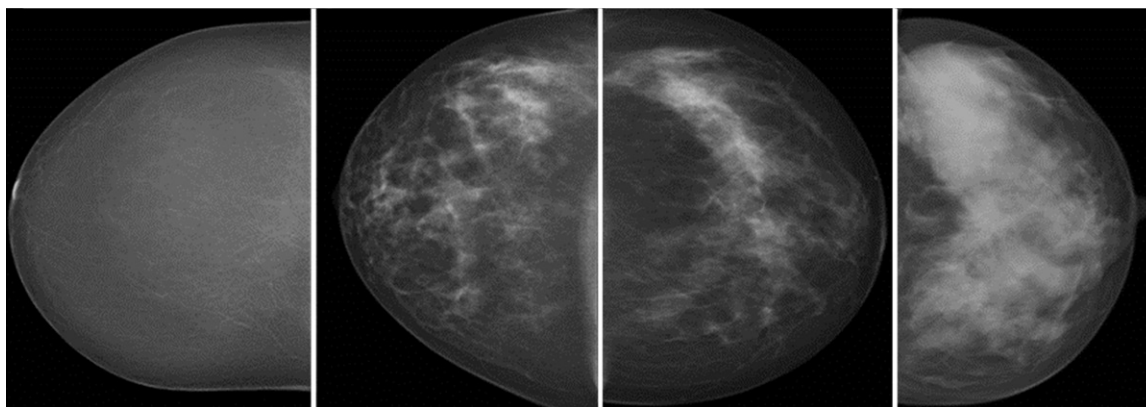
Figure 2: Mammograms show volumetric breast density classified as Volpara density grade 1, 2, 3, and 4 (<4.5%, 4.5%–7.49%, 7.5%–15.49%, and ≥15.5%, respectively) with automated volumetric breast density measurement.

Table 4: Breast Cancer Risk Factor Characteristics Associated with Screening Examinations Performed in BreastScreen Norway by VBD

Parameter	Nondense (VBD <7.5%)	Dense (VBD ≥7.5%)	P Value	All (VBD <52%)
No. of examinations	126 637	54 519	…	181 156
Age at screening (y)*	59.4 ± 5.5	57.1 ± 5.4	<.001	58.7 ± 5.6
Screening location				
Rogaland	48.9	57.3	<.001	51.5
Hordaland	51.1	42.7	<.001	48.5
Body mass index (kg/m²)*	26.7 ± 4.2	23.1 ± 3.0	<.001	25.6 ± 4.2
No. of pregnancies lasting >6 mo				
None	32.3	33.4	<.001	32.6
One or two	33.0	37.0	<.001	34.2
Three or more	34.7	29.6	<.001	33.2
Use of hormonal therapy				
Never use	52.6	55.9	<.0001	53.6
Current use of estrogen-progestin therapy[†]	4.1	6.9	<.0001	4.9
Current use of estrogen therapy[‡]	2.7	3.1	<.0001	2.8
Past use	23.0	18.7	<.0001	21.7
Unspecified[§]	17.6	15.4	<.0001	17.0
Positive family history[‖]	16.7	18.4	<.0001	17.2

Note.—Unless otherwise specified, data are percentages. VBD = volumetric breast density.

* Data are means ± standard deviations.

[†] Hormonal therapy containing combined estrogen-progestin (estradiol and norethisterone acetate).

[‡] Hormonal therapy containing estrogen alone (estradiol and estriol, and others [tibolone, with estrogenic, progestogenic and weak androgenic activity]).

[§] Defined as current user of unspecified hormonal therapy or incomplete information on past use.

[‖] Defined as mother, sister, daughter, and/or grandmother diagnosed with breast cancer.

The adjusted odds of a screen-detected breast cancer were 1.37 times higher (95% confidence interval: 1.19, 1.59) for screening examinations of women with dense versus nondense breasts (Table 5). Compared with women with nondense breasts, women with dense breasts had 2.93 times higher (95% confidence interval: 2.16, 3.97) odds of an interval breast cancer.

By using a continuous volumetric density measure, the predicted number of screen-detected breast cancers per 1000 screening examinations increased monotonically with increasing breast density and was highest for women aged at least 65 years (Fig 3). Women with very low breast density (VBD <4%) had the lowest predicted number of cases of screen-detected breast cancer (less than five per 1000 screening examinations). Screening examinations of women aged at least 65 years with very high breast density (about 30% or higher) had the highest predicted number of screen-detected breast cancers (approximately 13 per 1000 screening examinations). The predicted numbers of interval breast cancers were approximately one per 1000 screening examinations for those with VBD less than 4%, increasing to four per 1000 screening examinations for those with breast density greater than or equal to 10%, irrespective of age.

Table 5: Adjusted Odds Ratios and 95% Confidence Intervals of Screen-detected and Interval Breast Cancers in BreastScreen Norway

Parameter	Adjusted Odds Ratios of Screen-detected Breast Cancer[*]			Adjusted Odds Ratios of Interval Breast Cancer[†]		
	Odds Ratio	95% Confidence Interval	P Value	Odds Ratio	95% Confidence Interval	P Value
VBD						
Nondense (VBD <7.5)	1.00	…	…	1.00	…	…
Dense (VBD ≥7.5)	1.37	(1.19, 1.59)	<.0001	2.93	(2.16, 3.97)	<.0001
Age (y)	1.05	(1.04, 1.07)	<.0001	1.04	(1.01, 1.07)	.018
Screening location						
Rogaland	1.00	…	…	1.00	…	…
Hordaland	0.96	(0.85, 1.09)	.557	0.86	(0.66, 1.12)	.257
Screening history						
Prevalent	2.22	(1.81, 2.73)	<.0001	1.32	(0.87, 2.00)	.199
Regular, subsequent	1.00	…	…	1.00	…	…
Irregular, subsequent	2.01	(1.56, 2.60)	<.0001	1.18	(0.58, 2.39)	.648
Body mass index (kg/m²)						
Underweight (<18.5)	0.73	(0.53, 1.01)	.061	0.55	(0.28, 1.10)	.091
Normal (18.5–24.9)	1.00	…	…	1.00	…	…
Overweight (25.0–29.9)	1.12	(0.96, 1.29)	.146	1.32	(0.97, 1.80)	.081
Obese (≥30)	1.49	(1.24, 1.80)	<.0001	1.59	(1.01, 2.50)	.044
Use of hormonal therapy						
No use	1.00	…	…	1.00	…	…
Current use of estrogen-progestogen therapy[‡]	1.87	(1.49, 2.34)	<.0001	2.36	(1.54, 3.63)	<.0001
Current use of estrogen therapy[§]	1.35	(0.97, 1.88)	.074	2.09	(1.16, 3.76)	.014
Past use	0.89	(0.75, 1.06)	.181	1.02	(0.71, 1.47)	.925
Unspecified[‖]	0.84	(0.70, 1.02)	.076	1.04	(0.70, 1.56)	.833
No. of pregnancies lasting >6 mo						
None	1.00	…	…	1.00	…	
One or two	0.95	(0.81, 1.11)	.519	1.19	(0.87, 1.64)	.277
Three or more	0.78	(0.67, 0.92)	.003	0.84	(0.59, 1.19)	.326
Positive family history[#]						
No	1.00	…	…	1.00	…	…
Yes	1.49	(1.28, 1.75)	<.0001	1.86	(1.39, 2.50)	<.0001

Note.—Data were adjusted for age at screening, screening location, and screening history. VBD = volumetric breast density.

[*] 181 156 screening examinations from 2007–2015.

[†] 134 379 screening examinations from 2007–2013 with 2-year follow-up.

[‡] Hormonal therapy containing combined estrogen-progestin (estradiol and norethisterone acetate).

[§] Hormonal therapy containing estrogen alone (estradiol and estriol, and others [tibolone, with estrogenic, progestogenic and weak androgenic activity]).

[‖] Defined as current user of unspecified hormonal therapy or incomplete information on past use.

[#] Defined as mother, sister, daughter, and/or grandmother diagnosed with breast cancer.

Discussion

By using automated volumetric mammographic density assessment software, we identified higher rates of recall and biopsy and higher odds of screen-detected and interval breast cancer for screening examinations of women with volumetrically dense versus nondense breasts. Overall, 28% of the screening examinations in our study population were classified as volumetrically dense (VDG3 or VDG4), whereas 5% were categorized as very dense (VDG4). Tumor diameter and lymph node involvement of screen-detected cancers were less favorable for women with volumetrically dense versus nondense breasts.

The higher recall and biopsy rates for screening examinations of women with dense versus nondense breasts were expected (7,25) and likely related to challenges in interpretation (4). However, the highest recall rate in our study was substantially lower than recall rates in the United States, which are usually reported to be about 10% (26). The lower recall rates in Europe, including Norway, are likely related to independent double reading of mammograms with consensus for equivocal cases. It should be mentioned that a relatively small difference in biopsy rates for examinations with dense (1.4%) versus nondense (1.1%) breasts is clinically relevant for a screening program that covers about 85%

of the 600 000 women in the target population. The lower sensitivity of mammography screening among women with dense breasts indicates that mammographic screening is less effective for these women. This is because of a higher rate of interval breast cancers with less favorable tumor characteristics compared with screen-detected breast cancers. Our results support findings from studies based on subjectively assessed mammographic density and associated screening outcomes (25,27). A recent study from the Netherlands, using the same software for density assessment that we used, showed comparable results to our study among women with very high (VDG4) versus very low density (VDG1) (7).

Our findings of larger tumor diameter and a higher proportion of lymph node–positive tumors for screen-detected breast cancers among women with dense breasts are in line with other studies (11,28). The higher aggressiveness of tumors among women with dense breasts might indicate a combination of biologic characteristics and masking effect. The differences observed between women with nondense and dense breasts did not reach statistical significance for interval breast cancers, possibly because of the relatively small number of examinations categorized as dense. We assume the lower proportion of ductal carcinoma in situ for nondense versus dense breasts to be also related to biologic differences. However, further studies exploring mammographic features of ductal carcinoma in situ are needed to understand this finding.

Our results of higher odds of both screen-detected and interval breast cancers for women with dense versus nondense breasts (1.37 and 2.93, respectively) are in line with results from other studies (7,29–31). However, the predicted number of screen-detected and interval breast cancers per 1000 screening examinations were at most five for examinations with nondense breasts and were 13 for those with dense breasts. Thus, although the odds of breast cancer were higher for those with dense versus nondense breasts, the absolute risk of breast cancer was low. The predicted number would be substantially lower for VDG4, which corresponds to only 5% of the screening examinations.

The distribution of BI-RADS density into categories (almost entirely fatty, scattered fibroglandular densities, heterogeneously dense, and extremely dense) is reported to be approximately 10%, 40%, 40%, and 10%, respectively, for U.S. women in all ages (13). Correspondingly, the distribution of volumetric density in our sample was 38% (116 037 of 307 015), 35% (103 957 of 307 015), 24% (72 796 of 307 015), and 5% (14 225 of 307 015) for VDG1, VDG2, VDG3, and VDG4, respectively. The percentage of examinations classified into the two highest densities (VDG3 or VDG4) was 28%, which is lower than the expected distribution according to BI-RADS. However, the distribution of our study population corresponded well with another European study (7). Our study included information about screening examinations performed
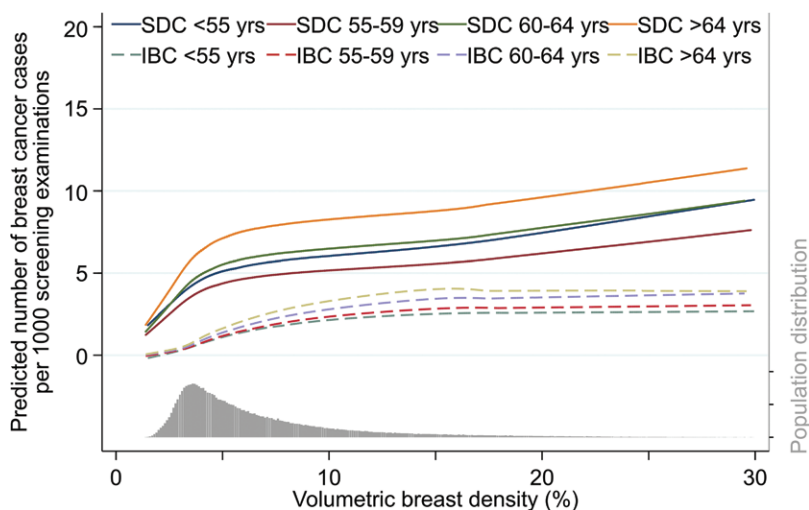
Figure 3: Graph shows smoothed predicted number of cases per 1000 screening examinations for screen-detected breast cancer (SDC) and interval breast cancer (IBC) in BreastScreen Norway by using volumetric breast density stratified by 5-year age groups. Models for risk of SDC and IBC included cubic spline function with six knots. Covariates included continuous volumetric breast density, age at screening (continuous), screening location and history, body mass index, number of pregnancies longer than 6 months, first- and second-degree family history of breast cancer, and use of hormonal therapy.

among women aged 50–69 years who participated in a nationwide program offering biennial mammographic screening, which is different from the United States where women usually are screened annually from age 40 years.

To our knowledge, our study is one of the largest to date on volumetric mammographic density and performance measures in an organized breast cancer screening program, including 1791 cases of screen-detected cancer and 384 cases of interval breast cancer. Furthermore, to assess the odds of breast cancer associated with volumetric mammographic density categorization, we obtained standardized data on breast cancer risk factors from a patient questionnaire with a high rate of response (23). The use of GE digital mammography systems only provided consistency in measurement.

Our study had some limitations. Despite the high completeness of data, we had some missing values for tumor characteristics and risk factors. Women included in the nondense group differed in some characteristics other than breast density from those in the dense group, making it difficult to conclude that differences in outcomes were solely because of breast density. To mitigate this issue, we included propensity scores in our analyses. Despite this correction, the results remained stable. The Bonferroni correction reduced the level of significance for each statistical test from the standard level, which limited the amount of statistically significant findings.

Furthermore, VBD was determined by using nonprocessed images. Mammographic density might be different when assessed by the radiologist or any other software by using processed images (12,30). The correlation of BI-RADS density categorization between radiologists in the United States and Norway is not known. The women with dense and nondense breasts do not share the same distributions for the available covariates. Thus,

this study had limitations related to the statistical uncertainty when constructing comparable groups.

In summary, by applying automated volumetric density measurements to our population-based screening program for women aged 50–69 years, we found that screening examinations of women having dense breasts showed higher rates of recall and biopsy, and higher odds of screen-detected and interval breast cancers than women with nondense breasts. Our results can be used to help inform how automated volumetric density categorization will change population-based screening performance and outcomes under a more objective breast density measurement paradigm.

References

1. Lee CI, Chen LE, Elmore JG. Risk-based breast cancer screening: implications of breast density. Med Clin North Am 2017;101(4):725–741.
2. Trentham-Dietz A, Kerlikowske K, Stout NK, et al. Tailoring breast cancer screening intervals by breast density and risk for women aged 50 years or older: collaborative modeling of screening outcomes. Ann Intern Med 2016;165(10):700–712.
3. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med 2007;356(3):227–236.
4. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 2000;92(13):1081–1087.
5. 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.
6. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996;276(1):33–38.
7. Wanders JO, Holland K, Veldhuis WB, et al. Volumetric breast density affects performance of digital screening mammography. Breast Cancer Res Treat 2017;162(1):95–103.
8. Sherratt MJ, McConnell JC, Streuli CH. Raised mammographic density: causative mechanisms and biological consequences. Breast Cancer Res 2016;18(1):45.
9. Berg WA, Zhang Z, Lehrer D, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA 2012;307(13):1394–1404.
10. Moshina N, Ursin G, Hoff SR, et al. Mammographic density and histopathologic characteristics of screen-detected tumors in the Norwegian Breast Cancer Screening Program. Acta Radiol Open 2015;4(9):2058460115604340.
11. Bertrand KA, Tamimi RM, Scott CG, et al. Mammographic density and risk of breast cancer by age and tumor characteristics. Breast Cancer Res 2013;15(6):R104.
12. Eng A, Gallant Z, Shepherd J, et al. Digital mammographic density and breast cancer risk: a case-control study of six alternative density assessment methods. Breast Cancer Res 2014;16(5):439.
13. Sickles E, 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; 123–126.
14. Singh JM, Fallenberg EM, Diekmann F, et al. Volumetric breast density assessment: reproducibility in serial examinations and comparison with visual assessment. Rofo 2013;185(9):844–848.
15. Desreux J, Bleret V, Lifrange E. Should we individualize breast cancer screening? Maturitas 2012;73(3):202–205.
16. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. Ann Intern Med 2011;155(1):10–20.
17. DenseBreast-info. Legislation and Regulations 2017. http://densebreast-info.org/legislation.aspx. Published December 15, 2017. Accessed November 15, 2017.
18. European Commission Initiative on Breast Cancer: new European guidelines for breast cancer screening. http://ecibc.jrc.ec.europa.eu/. Accessed February 21, 2018.
19. Larsen IK, Småstuen M, Johannesen TB, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. Eur J Cancer 2009;45(7):1218–1231.
20. 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. Accessed February 21, 2018.
21. 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.
22. 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.
23. 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 Jan 1:1403494817749393 [Epub ahead of print].
24. Goldhirsch A, Winer EP, Coates AS, et al. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 2013;24(9):2206–2223.
25. van Gils CH, Otten JD, Verbeek AL, Hendriks JH, Holland R. Effect of mammographic breast density on breast cancer screening performance: a study in Nijmegen, the Netherlands. J Epidemiol Community Health 1998;52(4):267–271.
26. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med 2011;155(8):493–502.
27. Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology 2008;246(2):376–383.
28. Eriksson L, Czene K, Rosenberg L, Humphreys K, Hall P. The influence of mammographic density on breast tumor characteristics. Breast Cancer Res Treat 2012;134(2):859–866.
29. Aiello EJ, Buist DS, White E, Porter PL. Association between mammographic breast density and breast cancer tumor characteristics. Cancer Epidemiol Biomarkers Prev 2005;14(3):662–668.
30. Brandt KR, Scott CG, Ma L, et al. Comparison of clinical and automated breast density measurements: implications for risk prediction and supplemental screening. Radiology 2016;279(3):710–719.
31. Wanders JOP, Holland K, Karssemeijer N, et al. The effect of volumetric breast density on the risk of screen-detected and interval breast cancers: a cohort study. Breast Cancer Res 2017;19(1):67.