Digital Breast Tomosynthesis with Synthesized Two-Dimensional Images versus Full-Field Digital Mammography for Population Screening: Outcomes from the Verona Screening Program[1]

Francesca Caumo, MD
Manuel Zorzi, MD
Silvia Brunelli, MD
Giovanna Romanucci, MD
Rossella Rella, MD
Loredana Cugola, MD
Paola Bricolo, MD
Chiara Fedato, MD
Stefania Montemezzi, MD
Nehmat Houssami, MD, PhD

Purpose: To examine the outcomes of a breast cancer screening program based on digital breast tomosynthesis (DBT) plus synthesized two-dimensional (2D) mammography compared with those after full-field digital mammography (FFDM).

Materials and Methods: This prospective study included 16666 asymptomatic women aged 50–69 years who were recruited in April 2015 through March 2016 for DBT plus synthetic 2D screening in the Verona screening program. A comparison cohort of women screened with FFDM (n = 14423) in the previous year was included. Screening detection measures for the two groups were compared by calculating the proportions associated with each outcome, and the relative rates (RRs) were estimated with multivariate logistic regression.

Results: Cancer detection rate (CDR) for DBT plus synthetic 2D imaging was 9.30 per 1000 screening examinations versus 5.41 per 1000 screening examinations with FFDM (RR, 1.72; 95% confidence interval [CI]: 1.30, 2.29). CDR was significantly higher in patients screened with DBT plus synthetic 2D imaging than in those screened with FFDM among women classified as having low breast density (RR, 1.53; 95% CI: 1.13, 2.10) or high breast density (RR, 2.86; 95% CI: 1.42, 6.25). The positive predictive value (PPV) for recall was almost doubled with DBT plus synthetic 2D imaging: 23.3% versus 12.9% of recalled patients who were screened with FFDM (RR, 1.81; 95% CI: 1.34, 2.47). The recall rate was similar between groups (RR, 0.95; 95% CI: 0.84, 1.06), whereas the recall rate with invasive assessment was higher for DBT plus synthetic 2D imaging than for FFDM (RR, 1.93; 95% CI: 1.31, 2.03). The mean number of screening studies interpreted per hour was significantly lower for screening examinations performed with DBT plus synthetic 2D imaging (38.5 screens per hour) than with FFDM (60 screens per hour) ($P < .001$).

Conclusion: DBT plus synthetic 2D imaging increases CDRs with recall rates comparable to those of FFDM. DBT plus synthetic 2D imaging increased image reading time and the time needed for invasive assessments.

[1] From the UOSD Breast Unit ULSS9, Ospedale di Marzana, Piazzale Lambranzi, 1, 37034 Verona, Italy (F.C., S.B., G.R., L.C., P.B.); Veneto Tumour Registry, Veneto Region, Padua, Italy (M.Z.); Dipartimento di Scienze Radiologiche, Università Cattolica del Sacro Cuore, Rome, Italy (R.R); Organizational Unit Prevention and Public Health, Veneto Region, Venice, Italy (C.F.); DAI Patologia e Diagnostica, Azienda Ospedaliera Universitaria Integrata, Verona, Italy (S.M.); and Sydney School of Public Health, Sydney Medical School, University of Sydney, Sydney, Australia (N.H.). Received March 30, 2017; revision requested June 6; revision received July 11; accepted August 8; final version accepted September 29. **Address correspondence to** F.C. (e-mail: *fcaumo@aulss9.veneto.it*).

N.H. supported by the National Breast Cancer Foundation Australia through a Breast Cancer Research Leadership Fellowship grant.

Several studies have shown the increased sensitivity and specificity of integrated two-dimensional (2D) imaging and digital breast tomosynthesis (DBT) in breast cancer population screening (1–5). Moreover, the use of two-view DBT is associated with substantially improved breast cancer detection rates (CDRs) (3,4,6,7). Synthesized 2D images were developed to reduce the radiation dose to the breast, a very important objective, especially for the screening of asymptomatic women. According to Skaane et al (8), the average glandular dose for full-field digital mammography (FFDM) plus DBT (3.52 mGy ± 1.08) was approximately double that of FFDM alone (1.58 mGy ± 0.61); synthetic 2D images, reconstructed from the DBT data by using specific software, reduced the radiation dose to acceptable values (1.95 mGy ± 0.58) for screening.

Very few studies have examined the performance of DBT with synthetic 2D imaging (DBT plus synthetic 2D imaging) in comparison with FFDM in real-world screening practice (9,10), and this screening strategy has not been validated in randomized controlled trials (10). Zuckerman et al (9) suggested that DBT plus synthetic 2D imaging has screening outcomes similar to those of DBT plus FFDM but with a significantly lower radiation dose. Bernardi et al (10) reported that DBT plus synthetic 2D imaging or DBT plus FFDM enabled detection of significantly more breast cancers than FFDM but increased the percentage of false-positive recalls.

We report the outcomes of a breast cancer screening program based on DBT plus synthetic 2D compared with FFDM screening.

Advances in Knowledge

- Screening with DBT plus synthesized 2D mammography increased cancer detection rate vs FFDM (9.30 vs 5.41 cancers per 1000 screens, $P < .001$) with comparable average glandular dose.
- Cancer stage distribution shows a lower proportion of in situ malignancy (9% with DBT plus synthetic 2D imaging vs 26.9% with FFDM) and higher proportion of stage I invasive cancer (72.3% with DBT plus synthetic 2D vs 50.0% with FFDM; $P = .11$) among patients whose cancer was detected with DBT plus synthetic 2D imaging versus FFDM.
- Recall rates for assessment were similar between DBT plus synthetic 2D and FFDM (4.0% vs 4.2%, respectively); however, we observed a significant increase in invasive assessment rate (1.5% vs 0.9%; $P < .001$).
- DBT plus synthetic 2D screening increased cancer detection rate in most age groups and density categories, with improved visualization of masses in high-density breasts (26 with DBT plus synthetic 2D vs five with FFDM) and architectural distortions in low-density breasts (12 with DBT plus synthetic 2D vs three with FFDM), respectively.
- Mean reading time doubled for DBT plus synthetic 2D imaging compared with FFDM as indicated by number of screenings interpreted per hour (38.5 vs 60 screens per hour; $P < .001$); percentage of discordant referral recommendations for cancers in double-reading practice decreased from 28.2% for FFDM to 7.1% for DBT plus synthetic 2D ($P = .0002$).

Implications for Patient Care

- DBT plus synthetic 2D imaging in population-based breast cancer screening programs improves cancer detection with evidence that this is generally consistent across age groups and breast density categories.
- DBT enables radiologists to increase detection of suspicious findings that require further investigation, with a higher cancer yield at assessment.
- Increased image reading time from DBT plus synthetic 2D imaging could be addressed by reconsidering whether double reading (standard practice in Europe) is necessary, given that DBT plus synthetic 2D imaging significantly reduced discordant recall for cancers compared with FFDM.

Materials and Methods

Study Setup and Patients

The study was granted institutional ethics approval. From April 1, 2015 to March 31, 2016, we prospectively recruited asymptomatic women (age range, 50–69 years) who participated in the Breast Cancer Screening Program in Verona, Italy. In Italy, the Breast Cancer Screening Program invites asymptomatic women (age range, 50–69 years) by letter to undergo screening with FFDM and provides population-based biennial breast screening in accordance with European screening standards (11). Screening examinations include

https://doi.org/10.1148/radiol.2017170745

Content codes: BR OI

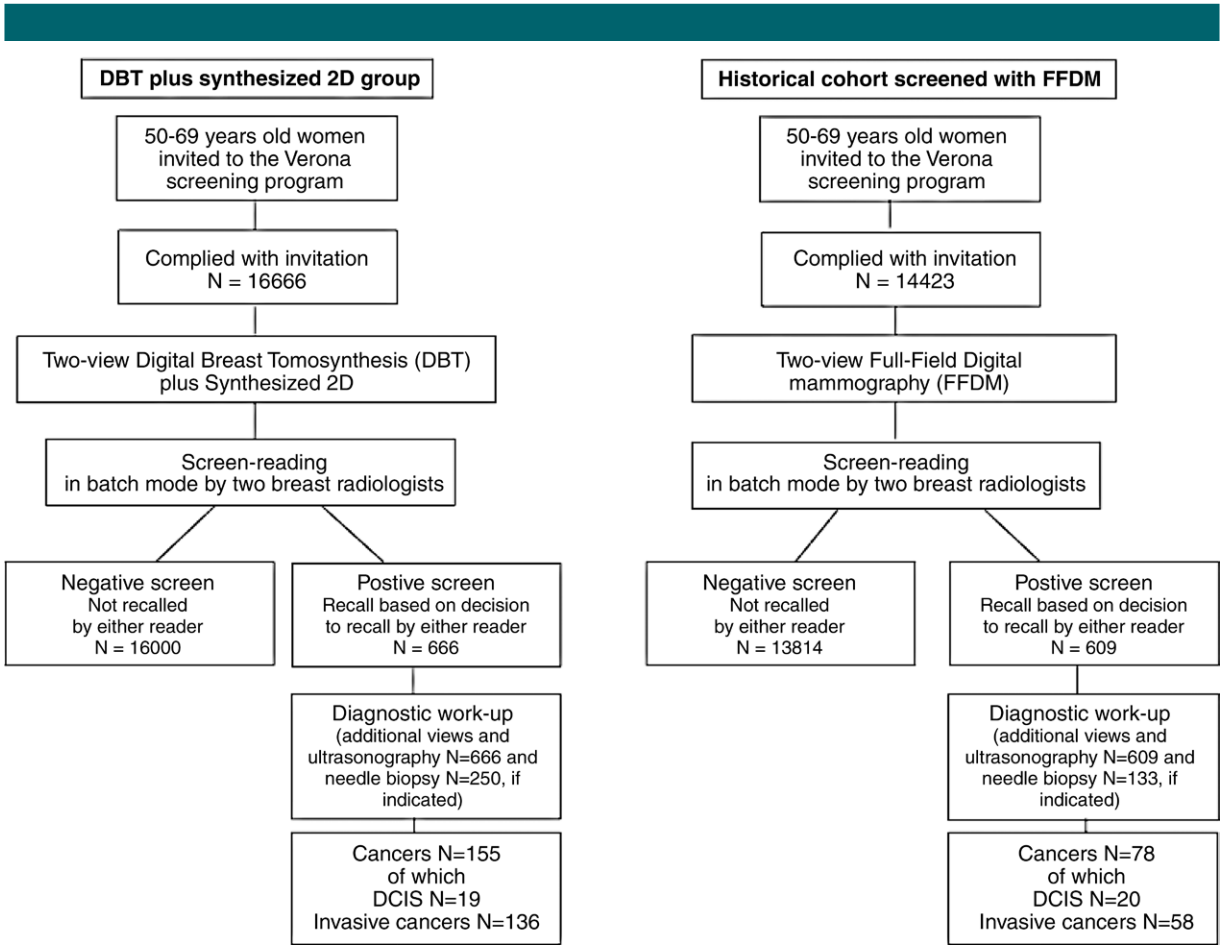
Radiology 2018; 287:37–46

Abbreviations:
BI-RADS = Breast Imaging Reporting and Data System
CDR = cancer detection rate
CI = confidence interval
DBT = digital breast tomosynthesis
DCIS = ductal carcinoma in situ
FFDM = full-field digital mammography
PPV = positive predictive value
RR = relative rate
2D = two-dimensional
VAB = vacuum-assisted biopsy

Author contributions:
Guarantors of integrity of entire study, F.C., S.B., L.C., P.B.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, F.C., S.B., G.R., R.R., P.B., S.M., N.H.; clinical studies, F.C., S.B., G.R., L.C., P.B., C.F., S.M.; statistical analysis, F.C., M.Z., S.B., G.R., L.C., P.B., C.F., S.M.; and manuscript editing, F.C., M.Z., S.B., G.R., R.R., P.B., S.M., N.H.

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

See also the editorial by Pisano in this issue.



Flowchart shows study design and subject flow.

two-view (craniocaudal and mediolateral oblique) FFDM of each breast. Each screening mammogram is independently interpreted, in batch mode, by two radiologists. A screening result is considered positive and the woman is recalled for further investigations if at least one reader records a positive finding. Diagnostic work-up for recalled women, referred to as second-level assessment, could potentially include additional DBT plus synthetic 2D views (both for the group screened with DBT plus synthetic 2D mammography and the comparison group), ultrasonography (US), and needle biopsy, if indicated. If the second-level assessment findings are negative, women undergo screening FFDM after 2 years; if the second-level assessment findings are positive, women are sent to the Breast Unit.

Participants' risk for breast cancer corresponds to that of the low-risk portion of the population. Women with a personal history of breast cancer, those who are BRCA mutation carriers, and high-risk women do not undergo breast cancer screening through the program and are triaged to specialized services.

All eligible women were invited by letter to undergo screening with DBT plus synthetic 2D mammography (the letter included a careful explanation of this new technique) or, alternatively, FFDM (Figure). All women chose to undergo DBT. Written informed consent for participation in the prospective study was obtained before DBT examination.

To enable comparison of screening detection measures, we assembled a historical cohort of women screened with FFDM in the same screening program in the year prior to the prospective study period (April 1, 2014 to March 31, 2015). Women in the historical cohort underwent screening with only FFDM. From October 2011 to June 2012, within our screening program 3040 women had been screened with DBT (4), but they were excluded from the present study to avoid introducing baseline differences between the two study cohorts of the current study. All women in the historical cohort were invited by letter to undergo screening with FFDM and gave their written consent for data from their participation to be used in research projects.

Procedures

Participants underwent DBT with a mammography unit (Selenia Dimensions;

Hologic, Bedford, Mass). Screening examinations included two-view (craniocaudal and mediolateral oblique) DBT mammography of each breast. The average glandular dose for a single DBT view was 2.09 mGy ± 0.55 (standard deviation) (1.13–3.65 mGy). In the comparison group, women underwent FFDM with the same system (Selenia Dimensions unit). The average glandular dose for a single-view FFDM was 1.48 mGy ± 0.58 (0.52–3.13 mGy).

Screening examinations were read at dedicated workstations with high-spatial-resolution mammography monitors. Screening mammograms were interpreted sequentially by radiologists initially by using synthetic 2D imaging alone and then by using DBT at the same screen-reading session. Reporting information was automatically collected by the radiology information system. The total number of readings by each radiologist within a 1-hour period included in the radiology information system–generated report was used to calculate the mean reading time.

Images from each screening examination (DBT plus synthetic 2D mammography or FFDM) were independently interpreted, in batch mode, by two of four breast radiologists (F.C., S.B., P.B., L.C.) with 3–13 years of experience in mammographic screening, and all the radiologists had 5 years of clinical experience in DBT interpretation. Double reading was used as the standard practice in European breast-screening programs. No computer-aided detection was used.

A screening examination was considered positive and a woman recalled for further investigations if at least one reader recorded a positive result (no arbitration was performed). Screening examinations were classified according to Breast Imaging Reporting and Data System (BI-RADS) categories: incomplete (BI-RADS 0), negative (BI-RADS 1), benign finding (BI-RADS 2), probably benign finding (BI-RADS 3), suspicious abnormality (BI-RADS 4), highly suggestive of malignancy (BI-RADS 5) (12). Readers recorded a positive result for BI-RADS 0 and BI-RADS category 3 or higher.

At the time of image reading, previous screening mammograms were routinely examined when available. For each screening examination, breast density data were also collected according to BI-RADS classification (12), and then the images were classified into two groups: low-density breast tissue, which included BI-RADS categories A and B, and high-density breast tissue, which included BI-RADS categories C and D. If the two readers had discordant density assessments, the final density assessment was decided by a third reader.

In the study group, at the same image reading session, readers collected data regarding lesion visibility in both DBT plus synthetic 2D images and in images obtained with DBT alone. Lesion data collected included DBT and synthetic 2D imaging findings (mass, microcalcification, architectural distortion, mass plus microcalcification); size (in millimeters); lesion localization ("clock" position); lesion side (right or left); BI-RADS assessment categories (DBT BI-RADS and synthetic 2D BI-RADS), and BI-RADS density classification.

Diagnostic work-up for recalled women (second-level assessment) could potentially include additional DBT plus synthetic 2D views (for both the group screened with DBT plus synthetic 2D imaging and the comparison group), US, and a needle biopsy, if indicated. Lesions classified as BI-RADS 4 or 5 underwent a US-guided core-needle biopsy (with a 14-gauge needle) or a DBT-guided vacuum-assisted biopsy (VAB) (with a 9-gauge needle). For DBT-guided VAB on areas of architectural distortion, 24 biopsy specimens were obtained in two biopsy rounds (mean sample weight, 4 g). Histopathologic results were classified as B1 to B5, according to European Guidelines (11): B1, normal tissue, regardless of whether breast parenchymal structures are present; B2, benign abnormalities such as sclerosing adenosis, fibroadenomas, fibrocystic changes, intramammary lymph nodes, duct ectasia, or fat necrosis; B3, a heterogeneous group of lesions of unknown biologic potential (atypical intraductal epithelial proliferation, flat epithelial atypia, lobular intraepithelial neoplasia including both atypical lobular hyperplasia and lobular carcinoma in situ, classic type, with low or intermediate nuclear grade, radial scar or complex sclerosing lesion, papillary lesion, and "other entities" including fibroepithelial lesion with cellular stroma and "mucocele-like" lesion); B4, suspicious findings but insufficient for a definite diagnosis of malignancy; and B5, unequivocal malignancy. Category 5 includes four subcategories: B5a, in situ carcinomas; B5b, invasive carcinomas; B5c, invasive status nonassessable; and B5d, other malignancy).

Outcomes and Statistical Analysis

The following outcomes were calculated for the group screened with DBT plus synthetic 2D imaging and for the comparison group: *(a)* attendance rate (percentage of women invited); *(b)* rates of recall for assessment and invasive assessment (recalled women who underwent a biopsy; percentage of women screened); *(c)* CDR for breast cancer, both invasive and in situ, overall and for cancers larger than 20 mm in greatest dimension (pT2+; per 1000 screened); *(d)* positive predictive value (PPV) for recall—screening examinations that resulted in recall (an initial interpretation of abnormal findings), after which a diagnosis of breast cancer was made (percentage); *(e)* the number of biopsies yielding borderline (B3) histologic findings (percentage); *(f)* the number of cancers detected by only one reader from double-reads (percentage); *(g)* the mean number of examinations interpreted within a 1-hour time period (average reading time); and *(h)* the storage requirement per year and per number of examinations.

The results were stratified by age-group (5-year classes), breast density (low or high) (12), and screening mammography (first or subsequent). A χ^2 test was used to compare distribution by age and breast density among groups. We compared the results of the two screening strategies by calculating the relative rates (RRs) of the main screening detection measures,

with 95% confidence intervals (CIs), by using a multivariate logistic regression analysis that included age as a covariate. Analyses were performed with statistical software (Stata 14.1; Stata, College Station, Tex). $P < .05$ was considered to indicate a significant difference.

Results

A total of 16666 participants (median age, 59 years) were screened with DBT plus synthetic 2D imaging during the study period, and 14423 had been screened in the prior year with FFDM alone (median age, 58 years) (Table 1). The rates of response to the invitation to undergo screening were 75.0% for the DBT plus synthetic 2D imaging group and 77.9% for the comparison group (RR, 0.96; 95% CI: 0.93, 0.99; $P = .01$). Descriptive data for each group are shown in Table 1. There were 155 cancers (136 invasive and 19 ductal carcinoma in situ [DCIS]) detected in the DBT plus synthetic 2D imaging group versus 78 cancers (58 invasive and 20 DCIS) ($P = .54$) in those screened with FFDM alone. The corresponding CDRs were 9.30 per 1000 screening examinations with DBT plus synthetic 2D imaging versus 5.41 per 1000 screening examinations with FFDM alone (RR, 1.72; 95% CI: 1.30, 2.29). As shown in Table 2, significantly increased cancer detection for DBT plus synthetic 2D screening was evident for first screening and subsequent (repeat) screening. The rate of detection of large (pT2+) tumors was similar between the DBT plus synthetic 2D imaging and FFDM groups (1.18 per 1000 vs 1.14 per 1000, respectively).

Screening with DBT plus synthetic 2D imaging was associated with increased CDRs in patients in all age groups except those aged 50–54 years (Table 3). The percentage of participants with high-density breasts was similar in both groups (2783 of 16666 [16.7%] in the DBT group vs 2436 of 14423 [16.9%] in the FFDM group). The CDRs stratified by breast density are shown in Table 3; there was a significantly higher CDR in the DBT plus

Table 1

Recall and Detection Outcomes in Women Screened with DBT Plus Synthetic 2D Imaging or with FFDM Alone

Parameter	FFDM			DBT Plus Synthetic 2D Imaging			*P* Value		
	First	Subsequent	Total	First	Subsequent	Total	First	Subsequent	Total
Recall rate (percentage screened)	7.6 (150/1982)	3.7 (459/12441)	4.2 (609/14423)	6.4 (195/3027)	3.5 (471/13639)	4.0 (666/16666)	.15	.32	.34
Invasive assessment rate (percentage screened)	2.1 (41/1982)	0.7 (92/12441)	0.9 (133/14423)	2.4 (73/3027)	1.3 (137/13639)	1.5 (250/16666)	.44	<.001	<.001
Invasive assessment rate (percentage recalled)	27.5 (41/149)	20.2 (92/456)	22 (133/605)	37.8 (73/193)	37.7 (177/470)	37.7 (250/663)	.15	<.001	<.001
Detection rate (per 1000 screened)	8.07 (16/1982)	4.98 (62/12441)	5.41 (78/14423)	11.6 (35/3027)	8.8 (120/13639)	9.3 (155/16666)	.23	<.001	<.001
Detection rate of pT2+ (per 1000 screened)	2.52 (5/1982)	0.96 (12/12441)	1.18 (17/14423)	2.31 (7/3027)	0.88 (12/13639)	1.14 (19/16666)	.88	.82	.92
Proportion of pT2+ (percentage of cancers with available stage)	31.3 (5/16)	19.4 (12/62)	21.8 (17/78)	20.0 (7/35)	11.2 (12/107)	13.4 (19/142)	.42	.2	.15
PPV for recall (%)	10.7 (16/149)	13.5 (62/456)	12.9 (78/605)	18.1 (35/193)	25.5 (120/470)	23.3 (155/663)	.10	<.001	<.001

Note.—FFDM indicates findings in the historical comparison group from the Verona screening program. DBT plus synthetic 2D = prospectively recruited study group from the Verona screening program.

Table 2

RRs of Screening Outcomes in Women Screened with DBT Plus Synthetic 2D Imaging versus FFDM (Reference)

	First Mammography Screening		Subsequent (Repeat) Screening		Total	
Parameter	RR	95% CI	RR	95% CI	RR	95% CI
RR (percentage screened)	0.85	0.68, 1.07	0.94	0.82, 1.07	0.95	0.84, 1.06
Invasive assessment rate (percentage screened)	1.17	0.78, 1.76	1.75	1.35, 2.29	1.63	1.31, 2.03
Invasive assessment rate (percentage recalled)	1.37	0.87, 2.19	1.87	1.40, 2.51	1.72	1.34, 2.19
CDR (per 1000 screened)	1.43	0.77, 2.78	1.77	1.29, 2.44	1.72	1.30, 2.29
Detection rate of pT2+ (per 1000 screened)	0.92	0.25, 3.67	0.91	0.37, 2.22	0.97	0.48, 1.98
Proportion of pT2+ (percentage of cancers with available stage)	0.59	0.14, 2.75	0.57	0.22, 1.49	0.59	0.27, 1.29
PPV for recall (%)	1.69	0.87, 3.39	1.89	1.34, 2.68	1.81	1.34, 2.47

Note.—Data are overall outcomes and outcomes stratified by first and subsequent screenings.

synthetic 2D imaging group than in the FFDM group for both density strata.

The PPV for recall was 155 of 663 (23.3%) for DBT plus synthetic 2D imaging versus 78 of 605 (12.9%) with FFDM alone (RR, 1.81; 95% CI: 1.34, 2.47).

Of the cancers detected, DCIS comprised 12.3% with DBT plus synthetic 2D imaging (19 of 155) and 25.6% with FFDM alone (20 of 78) (P = .002). Of the cancers detected, invasive (stage I) cancers comprised 72.3% (112 of 155) in the DBT plus synthetic 2D imaging group and 50.0% (39 of 78) in the FFDM group (P = .11). The proportion of pT2+ cancers was 12.3% (19 of 155) versus 21.8% (17 of 78), respectively (P = .11). Mean tumor size was 15.1 mm for cancers detected with DBT plus synthetic 2D and 18.6 mm for cancers detected with FFDM alone (P = .052). Lymph node involvement was observed in 11.5% of patients who underwent FFDM alone (nine of 78) and in 26.9% who underwent DBT plus synthetic 2D imaging (39 of 145) (P = .032). Stage distribution of cancers detected with DBT plus synthetic 2D imaging was as follows: stage 0, 12.3% (19 of 155) versus 23.1% (18 of 78) screened with FFDM (P = .073); stage Ia, 63% (97 of 155) versus 50% (39 of 78) screened with FFDM (P = .34); stage Ib, 9.7% (15 of 155) versus 2.6% (two of 78) screened with FFDM (P = .064); stage IIa, 6.5% (10 of 155) versus 15.4% (12 of 78) screened with FFDM (P = .048); stage IIb, 6.5% (10 of 155) versus 3.8% (three of 78) screened with FFDM (P = .44); stage IIIa, 1.9% (three of 155) versus 3.8% (three of 78) screened with FFDM (P = .40); stage IIIc, 0.6% (one of 155) versus 0% (0 of 78) screened with FFDM (P = .48).

The recall rate was similar for the DBT plus synthetic 2D and comparison groups (666 of 16666 [4.0%] vs 609 of 14423 [4.2%], respectively; RR, 0.95; 95% CI: 0.84, 1.06) (Table 2). Recall rates did not significantly differ between DBT plus synthetic 2D imaging and comparison groups when examined by first or subsequent screening (195 of 3027 [6.4%] and 471 of 13639 [3.5%], respectively, for the DBT plus synthetic 2D imaging group compared with 150 of 1982 [7.6%] and 459 of 12441 [3.7%] for the FFDM group) (Table 2).

The invasive assessment rate differed significantly between the DBT plus synthetic 2D imaging group and FFDM group for first screening examination (73 of 3027 [2.4%] vs 41 of 1982 [2.1%]; RR, 1.17; 95% CI: 0.78, 1.76) and for subsequent screening examinations (177 of 13639 [1.3%] vs 92 of 12441 [0.7%]; RR, 1.75; 95% CI: 1.35, 2.29) (Table 2). Imaging features of target lesions influenced the biopsy method: The percentage of US-guided core-needle biopsies performed decreased slightly (from 74 of 133 [55.6%] to 125 of 250 [50.0%]; P = .56), whereas the number of DBT-guided VABs increased from 59 of 133 (44.4%) in the FFDM group to 125 of 250 (50.0%) in the DBT plus synthetic 2D imaging group (P = .53). DBT-guided VAB allowed targeting for 32 lesions (11 of which were malignant) that were identified with DBT alone.

The number of biopsies yielding borderline (B3) histologic findings was 9.9% higher in the DBT plus synthetic 2D imaging group compared with the FFDM group (53 of 250 [21.2%] vs 15 of 133 [11.3%] patients, respectively; P = .0001).

Table 4 shows the imaging and histologic findings of lesions that underwent invasive assessment and the number of cancers stratified by breast density and imaging features. In particular, 13 of 155 (8.4%) of the cancers appeared as architectural distortions at DBT (vs four of 78 [5.1%] in the control group; P = .40); 11 B5 tumors were not visible on synthetic 2D images and were identified with DBT alone.

The number of cancers detected by only one reader during double-reads was 11 in the DBT group and 22 in the comparison group; hence, discordant recall for verified cancers decreased substantially with screening with DBT plus synthetic 2D imaging, from 28.2% (22 of 78) to 7.1% (11 of 155) (P < .001).

The mean number of examinations interpreted within a 1-hour time period was 38.5 for DBT plus synthetic 2D imaging and 60 for FFDM alone (P < .001).

Finally, the storage requirement per year was 5383118 megabytes (MB)—almost 323 MB per examination for DBT plus synthetic 2D imaging—versus

Table 3

Detection Rates of Cancer according to Age and Breast Density in Women Screened with DBT Plus Synthetic 2D Imaging versus Those in Women Screened with FFDM Alone

	FFDM			DBT Plus Synthetic 2D Imaging			DBT Plus Synthetic 2D Imaging vs FFDM		
Characteristic	Screened	Cancer	Detection Rate (per 1000 Screened)	Screened	Cancer	Detection Rate (per 1000 Screened)	RR*	95% CI	*P* Value
Age (y)									
50–54	3144	20	6.4	5129	35	6.8	1.07	0.60, 1.96	.80
55–59	3485	12	3.4	3605	28	7.8	2.26	1.11, 4.88	.016
60–64	3760	23	6.1	4066	53	13.0	2.13	1.28, 3.65	.002
65–69	4034	23	5.7	3866	39	10.1	1.77	1.03, 3.11	.028
Breast density†									
A or B	11987	67	5.6	13883	119	8.6	1.53	1.13, 2.10	.005
C or D	2436	11	4.5	2783	36	12.9	2.86	1.42, 6.25	.001

Note.—Except where indicated, data are numbers of patients.

* Reference group = FFDM.

† American College of Radiology BI-RADS.

504888 MB, almost 0.35 MB per examination, for FFDM alone.

Discussion

In most studies, DBT has been used in combination with 2D mammography, with double acquisitions and an increased radiation dose to the breast (13–19). Use of synthesized 2D images allows an acceptable radiation dose, in line with the European Guidelines for Quality Assurance in Breast Screening (11). In the Verona screening program, we transitioned to DBT plus synthetic 2D imaging, increasing CDR and maintaining similar rates of recall for assessment.

We observed that DBT plus synthetic 2D mammography had a higher invasive assessment rate than FFDM alone owing to improved visualization of breast cancers and B3 lesions. The PPV for cancer at second-level assessment was significantly higher in the DBT plus synthetic 2D imaging group than in the FFDM group.

There was evidence of improved cancer detection with DBT plus synthetic 2D imaging across age groups, with significant differences for women older than 55 years, further highlighting the effectiveness of DBT in terms of cancer detection for women in this age group. DBT plus synthetic 2D imaging also significantly improved cancer detection in all density-stratified analyses. There was a higher proportion of pT1 invasive cancers in the DBT group than in the FFDM group.

Furthermore, the proportion of DCIS was lower for DBT plus synthetic 2D screening than that for FFDM screening, but there is no statistical difference between DCIS numbers. DBT plus synthetic 2D imaging allows evaluation of microcalcification morphologic characteristics and—better than FFDM—the distribution through the sections. Bernardi et al (10) showed a decrease in recalls secondary to calcification. On the other hand, DBT could have a lower sensitivity compared with FFDM because some microcalcification may be less conspicuous with DBT (20). The higher detection of earlier-stage cancers with DBT plus synthetic 2D imaging implies that this modality has the potential to improve screening benefit. Bernardi et al (10) also detected a relatively higher proportion of earlier-stage cancer at DBT plus synthetic 2D imaging or DBT plus FFDM compared with FFDM alone. Zuckerman et al (9), however, found no statistical difference in the numbers of cancers detected or the proportions of invasive cancer versus DCIS detected at DBT plus synthetic 2D imaging compared with DBT plus FFDM. There is a further improvement in cancer detection compared with STORM (9.3 cancers per 1000 screening examinations vs 8.1 cancers per 1000 screen examinations for integrated screening) (4). This outcome could be explained by the further experience with DBT.

In our study, the percentage of participants with high-density breasts was similar in both groups but appeared to be significantly lower in comparison with other studies (9,21). Furthermore, the percentage of high-density breasts in Italian women is lower than in U.S. women because the Mediterranean diet and lifestyle are associated with lower breast density (22).

In our study, the percentage of US-guided core-needle biopsies performed decreased slightly, whereas the number of DBT-guided VABs increased. DBT-guided VAB allowed targeting for 32 lesions (11 of which were malignant) that were identified with DBT alone. On the other hand, the number B3 lesions was 9.9% higher in the DBT plus synthetic 2D imaging group than in the FFDM group, which suggests that DBT could increase surgical procedures for B3 lesions.

The mean reading time increased significantly for DBT plus synthetic 2D imaging compared with FFDM alone, with a corresponding reduction in the number of examinations interpreted per hour. There are more section images to

Table 4

Imaging Manifestation of Lesions that Underwent Invasive Assessment according to Histologic Result and Breast Density in Women Screened with DBT Plus Synthetic 2D Imaging versus That in Women Screened with FFDM Alone

	FFDM				DBT Plus Synthetic 2D Imaging			
Imaging Manifestation	Mass	Microcalcifications	Architectural Distortion	Mass with Microcalcifications	Mass	Microcalcifications	Architectural Distortion	Mass with Microcalcifications
	US-guided Core-Needle Biopsy (n = 74)				US-guided Core-Needle Biopsy (n = 125)			
B category*								
B1	0	0	0	0	0	0	0	0
B2	19	0	0	1	13	0	0	0
B3	5	0	0	1	3	0	2	1
B4	0	0	0	0	0	0	0	0
B5	42	0	0	6	87	0	8	11
Total	66	0	0	8	103	0	10	12
	VAB (n = 59)				VAB (n = 125)			
B category*								
B1	0	0	0	0	0	0	0	0
B2	5	13	1	1	10	13	5	1
B3	1	4	4	0	3	7	37	0
B4	0	0	0	0	0	0	0	0
B5	5	18	0	7	19	21	5	4
Total	11	35	5	8	32	41	47	5
Breast density	B5 Category				B5 Category			
A or B	38	14	3	12	80 (6)†	17 (0)†	12 (2)†	10 (0)†
C or D	5	4	1	1	26 (3)†	4 (0)†	1 (0)†	5 (0)†
Total	43	18	4	13	106 (9)†	21 (0)†	13 (2)†	15 (0)†

* Histologic results reported with B1–B5 scale (where B5 represents breast cancer) according to the European Guidelines (11).

† Number of lesions identified only with DBT are in parentheses.

be read when DBT is used, and image loading times are slower (4,23). The increased reading workload could be balanced by a reduction in the number of second-level work-ups or, alternatively, by elimination of double reading (24).

The European guidelines (11) recommend double reading of screening mammograms. Our data showed a significant decrease in discordant referral recommendations for verified cancers for DBT plus synthetic 2D imaging compared with FFDM (RR, 0.25; 95% CI: 0.11, 0.58; P = .0002). Only 11 breast cancers (7.1%) were detected by one reader in the DBT group (vs 28.2% in the FFDM group). Our results suggest that the usefulness of double reading appears to be reduced with DBT plus synthetic 2D imaging and are aligned with the findings reported by Houssami et al (24). The use of computer-aided detection could further assist in identifying cancer and further reduce the need for double reading (25), although its impact on recall warrants further evaluation.

Another aspect to be addressed in DBT screening is picture archiving and communication system image archive implementation. One DBT view takes up the same amount of storage space as a CT scan (about 80 MB), a highly relevant issue to consider when one is planning or extending the infrastructure to enable high-volume DBT screening.

The main limitation of this study is that it was a nonrandomized, historically controlled study. Therefore, the differences in screening outcomes between the two study groups could be confounded by other factors and might not be entirely attributable to screening technique. Another limitation is the lack of availability of data on interval cancers (the 2-year interval has not passed yet); the collection of those data are ongoing. Finally, because of differences in screening methods (2-year intervals, double reading, low-risk population) in other countries' screening programs, not all of our results are generalizable. In fact, breast cancer screening practice differs among countries and settings, with notable differences between European-based programs (women aged 50 years and older, biennial screening with double screen reading) and US-based breast cancer screening (women aged 40 years and older, annual screening with single-reading practice). These differences contribute to potential variability in the outcomes of DBT screening, as described in a summary by Houssami et al (26) of recent evidence.

In conclusion, the transition to DBT plus synthetic 2D imaging in the Verona screening program was associated with increased breast CDRs compared with screening with FFDM, with comparable recall rates for these screening strategies. PPV at second-level assessment also improved after transition to DBT plus synthetic 2D screening. However, the time needed for DBT image reading and the number of invasive assessments increased. Thus, before implementation of a screening strategy based on DBT, careful resource and infrastructure planning is required and should be complemented by rigorous health economic evaluation to inform screening policy decisions.

Disclosures of Conflicts of Interest: F.C. disclosed no relevant relationships. **M.Z.** disclosed no relevant relationships. **S.B.** disclosed no relevant relationships. **G.R.** disclosed no relevant relationships. **R.R.** disclosed no relevant relationships. **L.C.** disclosed no relevant relationships. **P.B.** disclosed no relevant relationships. **C.F.** disclosed no relevant relationships. **S.M.** disclosed no relevant relationships. **N.H.** disclosed no relevant relationships.

References

1. Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. Breast 2013;22(2):101–108.
2. 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.
3. 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.
4. 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.
5. 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.
6. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution x-ray imaging observer study. Radiology 2012;262(3):788–796.
7. Rafferty EA, Park JM, Philpotts LE, et al. Diagnostic accuracy and recall rates for digital mammography and digital mammography combined with one-view and two-view tomosynthesis: results of an enriched reader study. AJR Am J Roentgenol 2014;202(2):273–281.
8. 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.
9. Zuckerman SP, Conant EF, Keller BM, et al. Implementation of synthesized two-dimensional mammography in a population-based digital breast tomosynthesis screening program. Radiology 2016;281(3):730–736.
10. 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.
11. Perry N, Broeders M, de Wolf C, Toernberg S, Holland R, von Karsa L, eds. European Guidelines for Quality Assurance in Breast Screening and Diagnosis. Luxembourg: Office for Official Publications of the European Communities, 2006; 221–256.
12. D'Orsi CJ, Sickles EA, Mendelson EB, et al. ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. Reston, VA: American College of Radiology, 2013.
13. Houssami N. Digital breast tomosynthesis (3D-mammography) screening: data and implications for population screening. Expert Rev Med Devices 2015;12(4):377–379.
14. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014;311(24):2499–2507.
15. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomo-

synthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology 2013;269(3):694–700.

16. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. J Natl Cancer Inst 2014;106(11):dju316.
17. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. AJR Am J Roentgenol 2014;203(3):687–693.
18. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. Radiology 2015;274(1):85–92.
19. Svahn TM, Houssami N, Sechopoulos I, Mattsson S. Review of radiation dose estimates in digital breast tomosynthesis relative to those in two-view full-field digital mammography. Breast 2015;24(2):93–99.
20. Tagliafico A, Mariscotti G, Durando M, et al. Characterisation of microcalcification clusters on 2D digital mammography (FFDM) and digital breast tomosynthesis (DBT): does DBT underestimate microcalcification clusters? Results of a multicentre study. Eur Radiol 2015;25(1):9–14.
21. 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.
22. Razzaghi H, Troester MA, Gierach GL, Olshan AF, Yankaskas BC, Millikan RC. Mammographic density and breast cancer risk in White and African American Women. Breast Cancer Res Treat 2012;135(2):571–580.
23. Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading—evidence to guide future screening strategies. Eur J Cancer 2014;50(10):1799–1807.
24. Houssami N, Bernardi D, Pellegrini M et al. Breast cancer detection using single-reading of breast tomosynthesis (3D-mammography) compared to double-reading of 2D-mammography: evidence from a population-based trial. Cancer Epidemiol 2017;47:94–99.
25. Morra L, Sacchetto D, Durando M, et al. Breast cancer: computer-aided detection with digital breast tomosynthesis. Radiology 2015;277(1):56–63.
26. Houssami N, Hunter K, Zackrisson S. Overview of tomosynthesis (3D mammography) for breast cancer screening. Breast Cancer Manag 2017 (in press).