Accuracy of Digital Breast Tomosynthesis for Depicting Breast Cancer Subgroups in a UK Retrospective Reading Study (TOMMY Trial)[1]

Fiona J. Gilbert, FRCR
Lorraine Tucker, DCR
Maureen G. C. Gillan, PhD
Paula Willsher, DCR
Julie Cooke, FRCR
Karen A. Duncan, FRCR
Michael J. Michell, FRCR
Hilary M. Dobson, FRCR
Yit Yoong Lim, FRCR
Tamara Suaris, FRCR
Susan M. Astley, PhD
Oliver Morrish, MSc
Kenneth C. Young, PhD
Stephen W. Duffy, MSc

Purpose: To compare the diagnostic performance of two-dimensional (2D) mammography, 2D mammography plus digital breast tomosynthesis (DBT), and synthetic 2D mammography plus DBT in depicting malignant radiographic features.

Materials and Methods: In this multicenter, multireader, retrospective reading study (the TOMMY trial), after written informed consent was obtained, 8869 women (age range, 29–85 years; mean, 56 years) were recruited from July 2011 to March 2013 in an ethically approved study. From these women, a reading dataset of 7060 cases was randomly allocated for independent blinded review of *(a)* 2D mammography images, *(b)* 2D mammography plus DBT images, and *(c)* synthetic 2D mammography plus DBT images. Reviewers had no access to results of previous examinations. Overall sensitivities and specificities were calculated for younger women and those with dense breasts.

Results: Overall sensitivity was 87% for 2D mammography, 89% for 2D mammography plus DBT, and 88% for synthetic 2D mammography plus DBT. The addition of DBT was associated with a 34% increase in the odds of depicting cancer (odds ratio [OR] = 1.34, $P = .06$); however, this level did not achieve significance. For patients aged 50–59 years old, sensitivity was significantly higher ($P = .01$) for 2D mammography plus DBT than it was for 2D mammography. For those with breast density of 50% or more, sensitivity was 86% for 2D mammography compared with 93% for 2D mammography plus DBT ($P = .03$). Specificity was 57% for 2D mammography, 70% for 2D mammography plus DBT, and 72% for synthetic 2D mammography plusmDBT. Specificity was significantly higher than 2D mammography ($P < .001$in both cases) and was observed for all subgroups ($P < .001$ for all cases).

Conclusion: The addition of DBT increased the sensitivity of 2D mammography in patients with dense breasts and the specificity of 2D mammography for all subgroups. The use of synthetic 2D DBT demonstrated performance similar to that of standard 2D mammography with DBT. DBT is of potential benefit to screening programs, particularly in younger women with dense breasts.

[1] From the Department of Radiology, University of Cambridge, Cambridge Biomedical Campus, Cambridge CB2 0QQ, England (F.J.G., L.T., P.W.); Department of Radiology, University of Aberdeen, Aberdeen, Scotland (M.G.C.G.); Jarvis Breast Centre, Guildford, England (J.C.); North East Scotland Breast Screening Centre, Aberdeen, Scotland (K.A.D.); Department of Radiology, King's College Hospital, London, England (M.J.M.); West of Scotland Breast Screening Service, Glasgow, Scotland (H.M.D.); Department of Radiology, University Hospital South Manchester, Manchester, England (Y.Y.L.); Department of Radiology, St Batholomew's Hospital, London, England (T.S.); Department of Imaging Science and Biomedical Engineering, University of Manchester, Manchester, England (S.M.A.); East Anglian Regional Radiation Protection Service, Cambridge University Hospitals, Cambridge, England (O.M.); National Coordinating Centre for Physics of Mammography, Royal Surrey County Hospital, Guildford, England (K.C.Y.); and Centre for Cancer Prevention, Queen Mary University of London, London, England (S.W.D.). From the 2013 RSNA Annual Meeting. Received November 4, 2014; revision requested December 13; revision received March 10, 2015; accepted March 20; final version accepted May 1. The TOMMY Trial (a comparison of digital breast tomosynthesis with mammography in the UK Breast Screening Programme) was supported by the NIHR Health Technology Assessment Programme. **Address correspondence to** F.J.G. (e-mail: *fjg28@medschl.cam.ac.uk*).

Breast screening with mammography is recognized as an effective method for detecting early-stage breast cancer (1). However, the presence of overlapping breast fibroglandular tissue can decrease the visibility of cancers or simulate the appearance of a malignant lesion, reducing sensitivity and increasing false-positive results (2). It has been reported that 15%–30% of cancers are not detected at standard screening; this percentage is higher in women younger than 50 years and with dense breasts (3–6).

The development of digital breast tomosynthesis (DBT) and its clinical application has been summarized in a number of review articles, and its performance either relative to or in combination with two-dimensional (2D) mammography in clinical studies has been reviewed (7–15). Most studies were small retrospective reader performance studies with cancer enriched datasets rather than screening studies, but they provided evidence that DBT has the potential to improve the accuracy of mammography by reducing screening recall rates and increasing cancer detection rates. More recently, prospective screening studies demonstrated improved sensitivity and specificity with the use of DBT. The Oslo Tomosynthesis Screening Trial reported a 27% increase in the cancer detection rate across all breast densities and a 15% decrease in the false-positive recall rate when DBT was used in combination with 2D mammography compared with 2D mammography alone (16). The population-based Screening with Tomosynthesis or Mammography (STORM) study compared sequential 2D mammography and combined DBT and 2D mammography and reported a 34% increase in cancer detection across all age groups and breast densities and a potential to reduce the false-positive recall rate by 17% (17).

Current evidence favors the use of two-view DBT as an adjunct to 2D mammography rather than as a stand-alone imaging modality, although this practice almost doubles the radiation dose (14). To address this concern, a synthetic 2D image can be created from the tomosynthesis data, a procedure that is currently being evaluated in the Oslo trial (18,19).

While the STORM and Oslo studies demonstrated the improved accuracy of 2D mammography plus DBT in screening, studying the performance of DBT in specific subgroups of patients requires that data be enriched with cancer cases. There is a need to establish whether the addition of DBT is equally effective in women with different breast densities, ages, and tumor subtypes and sizes. Our purpose was to compare the diagnostic performance of 2D mammography, 2D mammography plus DBT, and synthetic 2D mammography plus DBT for depicting malignant features in different subgroups of women invited for screening.

Advances in Knowledge

- The diagnostic accuracy of digital breast tomosynthesis (DBT) with either two-dimensional (2D) or synthetic 2D mammography was significantly improved compared with 2D mammography alone ($P < .001$ in both cases).
- The area under the receiver operating characteristics curve (AUC) was 0.84 (95% confidence interval [CI]: 0.83, 0.86) for 2D mammography alone, 0.89 (95% CI: 0.87, 0.90) for 2D mammography plus DBT, and 0.87 (95% CI: 0.86, 0.89) for synthetic 2D mammography plus DBT.
- The addition of DBT increased the sensitivity of 2D mammography in women with dense breasts (86% for 2D mammography alone vs 93% for 2D mammography plus DBT) and the specificity of 2D mammography in all subgroups (58% for 2D mammography alone vs 69% for 2D mammography plus DBT).
- Overall, the sensitivity and specificity of synthetic 2D mammography plus DBT (88% and 71%, respectively) were comparable to those of 2D mammography plus DBT (89% and 69%, respectively).
- A significant increase in sensitivity was observed for 2D mammography plus DBT ($P = .04$) when the dominant radiologic feature was a mass, with 89% (range, 86%–92%) sensitivity for 2D mammography and 92% (range, 89%–95%) for 2D mammography plus DBT.

Implications for Patient Care

- The use of DBT with either 2D or synthetic 2D mammography may be beneficial to screening programs, reducing the number of false-positive results.
- The addition of DBT is of particular benefit in younger women with dense breasts.
- Synthetic 2D mammography appears to have diagnostic accuracy similar to that of 2D mammography when used in conjunction with DBT.

Materials and Methods

Support for this study was provided by Hologic (Bedford, Mass), who created

Published online before print
10.1148/radiol.2015142566 **Content codes:** BR DM

Radiology 2016; 000:1–10

Abbreviations:
AUC = area under the ROC
CI = confidence interval
DBT = digital breast tomosynthesis
ROC = receiver operating characteristics curve
STORM = Screening with Tomosynthesis or Mammography
2D = two-dimensional

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

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

synthetic 2D images for the reading study. The authors had control of all data in the study and the information submitted for publication. The study was approved and given a favorable ethical opinion by Scotland A Research Ethics Committee.

Study Design and Participants

The study was designed to produce a cancer-rich cohort of cases so that comparisons of tumor size and characteristics could be analyzed. Women who were aged 47 years and older in six National Health Service breast screening program centers and recalled to an assessment clinic for a mammographic abnormality found at routine 2D screening and women who were younger than 50 years old with a family history of breast cancer and underwent annual mammography screening (family history cases) were eligible. No women underwent DBT before admittance to the study because it is not used in the United Kingdom screening program. Written informed consent was obtained. Women with breast implants, those who were pregnant, and those who were unable to give consent were excluded. Only one case per woman was included in the study.

Image Acquisition

At the assessment or family history clinic, all participants underwent standard two-view 2D mammography of both breasts and two-view DBT as a single procedure with the same breast compression and on the same digital mammography unit (Selenia Dimensions; Hologic, Bedford, Mass). Synthetic 2D images were generated from the DBT dataset for each case by using C-View 2011 image processing software (Hologic).

Readers

The 26 retrospective study readers comprised radiologists (n = 21), advanced practitioner radiographers (n = 3), and breast clinicians (n = 2), with an average of 10 years (range, 3–25) experience reading film in the National Health Service breast screening program. These readers were among a group of readers from the six centers who participated in the prospective data collection. Each reader attended a 1-day DBT reading course and read a test set of 80 cases. Readers also gained experience by reviewing DBT images acquired at their own site during recruitment. Images were viewed on SecurView DW workstations (Hologic) that were optimized to read both 2D and DBT images.

Prospective Data Collection

Assessment cases were read by a single reader with available 2D and DBT images, and family history cases were independently read by two readers with 2D and DBT images. Synthetic 2D images were not available during prospective data collection. For each case, readers marked the location and type of abnormalities, gave a five-point suspicion score, and assessed breast density by using a visual analog scale. For family history cases, an overall recall or no recall decision was also recorded.

Retrospective Reading Study

The planned reading target for this study was 7000 cases. The sample size calculation was powered to allow significant differences to be evaluated for subgroup analyses. The smallest anticipated subgroup of cancers was estimated to comprise around 15% of the total tumor population. We postulated that 2D mammography had sensitivity of 85% and that 2D mammography plus DBT had sensitivity of 95%. With a 5% significance level and two-sided testing, 90% power to detect this difference (ie, 5% missed by DBT and 15% missed by 2D mammography) as significant requires at least 38 cancers with discordant findings. Thus, 190 cancers (38 ÷ 0.2) were needed in the subgroup. As was previously stated, the smallest subgroup was expected to be 15% of the total; therefore, a total of 1267 cancer cases were required, which implies a total study size of 7000 cases. A study population of this size would have at least 90% power for any subgroup that constitutes at least 15% of the total study and 80% for any subgroup that constitutes at least 11% of the total study.

A randomization program managed by the Cambridge Clinical Trials Unit assigned weekly reading sets that included a mix of healthy, benign, and cancer cases. Readers reviewed either *(a)* 2D mammography, *(b)* 2D mammography plus DBT, or *(c)* synthetic 2D mammography plus DBT images for any one case (ie, individual cases were read by a different reader in each arm of the study) and randomized cases from all three arms of the study. Readers did not review any cases from their own center. Readers were blinded to the outcome status of each case, and cases were independently read without access to previous examinations. The decision to recall the case or not was made, and the location, size, and type of suspicious findings were recorded with a five-point suspicion score.

Outcome Measures

For cases who underwent biopsy, the outcomes from the three arms of the reading study were compared with the final histopathologic findings to verify the presence of benign or malignant disease. At assessment investigations, including a clinical examination, additional mammographic views and ultrasonographic scans were obtained as appropriate. Negative (benign or healthy) cases are routinely reviewed at the assessment clinic by two radiologists before the patient is discharged as a failsafe to ensure that nothing was overlooked.

Statistical Analysis

Data for all three reading arms for all cases were not available because images were not converted to synthetic 2D mammography, data were lost in transit to reading sites, and there were image management issues. The primary analysis included cases that were read in at least two arms of the study to avoid introducing bias. The analysis was rerun by using only cases that were read in all three arms and produced almost identical results. For the purposes of analysis, a study was categorized as

positive if the reader decided to recall the patient.

Sensitivities and specificities were calculated for each of the three reading arms: first for all cases, then for subgroups according to breast density and the dominant radiologic feature. In addition, sensitivity to cancers was calculated for subgroups by size, lymph node status, and histologic grade. Because each imaging modality was applied to the same cases, analysis of binary outcomes was carried out with McNemar methods and corresponding conditional logistic regression analyses (20,21). The methodology of the McNemar analysis implied that, for a comparison of two imaging modalities, only cases with no missing data for each modality were included. Statistical analysis was performed with Stata version 10.0 (StateCorp, College Station, Tex). We interpreted *P* values less than .05 as indicating significant difference.

Receiver operating characteristics (ROC) analysis was performed to compare diagnostic accuracy of the three study arms. Areas under the ROC curve (AUC) were compared by using the method set forth by De Long et al (22).

Results

As shown in Figure 1, a total of 8869 participants, aged 29–85 years (mean age, 56 years), were recruited, with 7684 assessment cases and 1185 family history cases. Exclusions (n = 207) resulted in a prospective dataset of 8662 cases. A further 1412 cases were randomly excluded to reduce the dataset in line with the planned reading target of 7000 cases. There were 190 cases for which, due to logistics and image transfer issues, data were available from only one reading arm. This resulted in a final dataset for analysis of 7060 cases (assessment, n = 6020; family history, n = 1040).

The characteristics of the study population are shown in Table 1. Reading data were available for 6927 (98%) 2D mammography cases, 6959 (99%) 2D mammography plus DBT cases, and 6653 (94%) synthetic 2D mammography plus DBT cases. Of the 26 readers, one read only 85 cases, and the others read between 493 and 1078 cases (mean, 851) that were evenly distributed over the three arms of the trial. The variation in the number of cases read was due to reader availability, the amount of storage space on workstations, and the randomization process (whereby no case could be read in the center from which it originated). Nonradiologists tended to have higher recall rates in all arms and regardless of cancer status (data not shown).

Although the primary aim of this study was to address specific subgroups, it is of illustrative value to see the overall matched comparison of 2D mammography with DBT. There were 1137 cancers and 5691 noncancers with reading data for both 2D mammography and 2D mammography plus DBT. In the cancer group, 921 cancers were depicted by both 2D mammography and 2D mammography plus DBT, 71 were depicted by 2D mammography alone, 95 were depicted by 2D mammography plus DBT alone, and 50 were not depicted by either modality. The addition of DBT conferred a 34% increase in the odds of detecting cancer (odds ratio, 1.34; 95% confidence interval [CI]: 0.97, 1.85; P = .06), a level that did not achieve significance, and a substantial improvement in specificity, with a 56% reduction in the odds of recalling noncancers (odds ratio, 0.44; 95% CI: 0.39, 0.49; $P < .001$) (Table 2).

Results of ROC analysis showed that the AUC was 0.84 (95% CI: 0.82, 0.86) for 2D mammography, 0.89 (95% CI: 0.87, 0.91) for 2D mammography plus DBT, and 0.88 (95% CI: 0.86, 0.90) for synthetic 2D mammography plus DBT (Fig 2). Both DBT arms had significantly greater AUCs than did 2D mammography ($P < .001$ in both cases).

For cases with breast density of 50% or more, AUCs were 0.83 (95% CI: 0.79, 0.86) for 2D mammography, 0.89 (95% CI: 0.87, 0.92) for 2D mammography plus DBT, and 0.87 (95% CI: 0.84, 0.90) for synthetic 2D mammography plus DBT. Both 2D mammography plus DBT and synthetic 2D mammography plus DBT had significantly greater AUCs than did 2D mammography ($P < .001$ in both cases). In cases where microcalcification was the dominant radiologic finding, AUCs were 0.73 (95% CI: 0.69, 0.77) for 2D mammography, 0.75 (95% CI: 0.71, 0.78) for 2D mammography plus DBT, and 0.74 (95% CI: 0.70, 0.78) for synthetic 2D mammography plus DBT. The AUCs of 2D mammography plus DBT (P = .5) and synthetic 2D mammography plus DBT (P = .5) did not significantly differ from that of 2D mammography. Differences in AUCs were mainly driven by specificity.

Figure 1

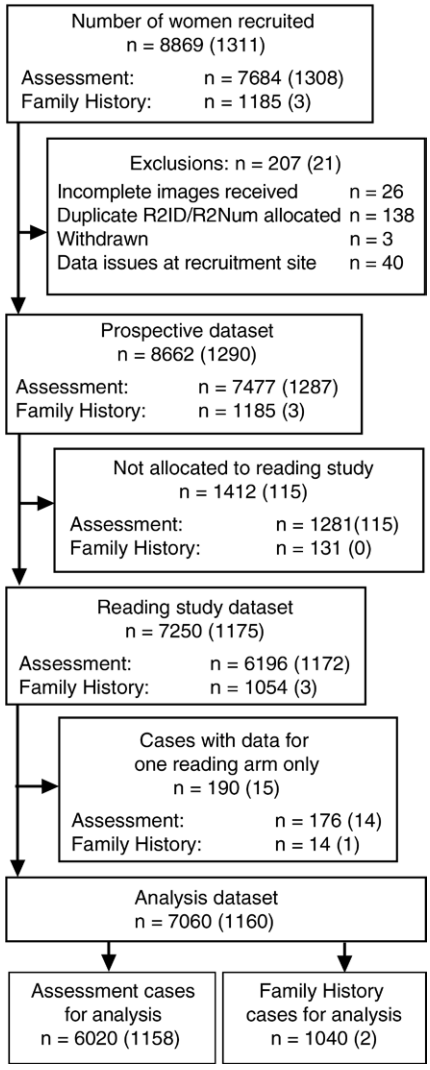

Figure 1: Flowchart shows the allocation of study subjects. The numbers in parentheses indicate the number of women with cancer.

Table 2 shows sensitivity and specificity for all subjects and specific subgroups. Overall, sensitivity was 87% (95% CI: 85, 89) for 2D mammography, 89% (95% CI: 87, 91) for 2D mammography plus DBT, and 88% (95% CI: 86, 90) for synthetic 2D mammography plus DBT. The difference in sensitivity between 2D mammography and 2D mammography plus DBT was of borderline significance (P = .06). There was no significant difference in sensitivity between 2D mammography and synthetic 2D mammography plus DBT (P = .6). Specificity was 58% (95% CI: 56, 60) for 2D mammography, 69% (95% CI: 67, 71) for 2D mammography plus DBT, and 71% (95% CI: 69, 73) for synthetic 2D mammography plus DBT. Specificity for 2D mammography plus DBT and synthetic 2D mammography plus DBT (P < .001 in both cases) was significantly higher than it was for 2D mammography. The increased specificity of 2D mammography plus DBT and synthetic 2D mammography plus DBT was observed in all subgroups of breast density, dominant radiologic feature, and age (P < .001 in all cases). For all three modalities, specificity was lower for depicting microcalcification and higher for depicting distortion/asymmetry, although a significant improvement in specificity for both DBT modalities was consistently observed in these categories.

For patients aged 50–59 years, sensitivity was significantly higher (P = .01) for 2D mammography plus DBT (91% [95% CI: 88, 94]) than it was for 2D mammography alone (87% [95% CI: 84, 90]), and for those with breast density of 50% or more, sensitivity was 86% (95% CI: 82, 90) for 2D mammography and 93% (95% CI: 90, 96) for 2D mammography plus DBT (P = .03). For depicting 11–20-mm invasive tumors, sensitivity was significantly higher (P < .001) for 2D mammography plus DBT (93% [95% CI: 90, 96]) than it was for 2D mammography (86% [95% CI: 82, 90]). In cases with grade 2 invasive tumors, sensitivity of 2D mammography plus DBT (91% [95% CI: 88, 94]) was increased compared to that of 2D mammography alone (87% [95%

Table 1

Participant Characteristics

	Assessment Cases		Family History Cases	
Characteristic	Randomized	Cancers	Randomized	Cancers
Age				
< 40 years	3 (<1)	1 (<1)	11 (1)	0
40–49 years	340 (6)	27 (2)	938 (94)	1 (50)
50–59 years	3568 (59)	462 (40)	44 (4)	1 (50)
60–69 years	1714 (29)	519 (45)	3 (< 1)	0
≥70 years	364 (6)	141 (12)	0	0
Unknown	31	9	44	0
Breast density				
0%–24%	1,636 (27)	378 (33)	233 (23)	0
25%–49%	2,556 (43)	439 (38)	418 (42)	1 (50)
50%–74%	1,376 (23)	271 (24)	271 (27)	1 (50)
75%–100%	396 (7)	63 (5)	83 (8)	0
Unknown	56	8	35	0
Cancer type				
Invasive ductal*	…	788 (68)	…	1 (50)
Invasive lobular*	…	109 (9)	…	
Invasive other*	…	59 (5)	…	
DCIS	…	203 (18)	…	1 (50)
Invasive cancer size				
1–5 mm	…	73 (8)	…	0
6–10 mm	…	243 (26)	…	0
11–20 mm	…	434 (46)	…	1 (100)
21–50 mm	…	183 (19)	…	0
> 50 mm	…	10 (1)	…	0
Unknown	…	13	…	0
DCIS size				
1–5 mm	…	30 (15)	…	0
6–10 mm	…	30 (15)	…	0
11–20 mm	…	47 (24)	…	0
21–50 mm	…	78 (39)	…	1 (100)
>50 mm	…	15 (7)	…	0
Unknown	…	3	…	0
Invasive cancer grade				
1	…	242 (26)	…	0
2	…	504 (54)	…	2 (100)
3	…	180 (20)	…	0
Unknown	…	30	…	0
DCIS grade				
Low	…	10 (8)	…	0
Intermediate	…	31 (22)	…	1 (100)
High	…	97 (70)	…	0
Unknown	…	65	…	0
Lymph node status†				
Normal	…	514 (58)	…	0
< four positive nodes	…	292 (33)	…	1 (100)
≥four positive nodes	…	77 (9)	…	0
Unknown	…	73	…	0
Dominant imaging feature				
Circumscribed mass	1814 (30)	145 (13)	84 (8)	0
Spiculated mass	712 (12)	508 (44)	3 (< 1)	0

Table 1 (continues)

Table 1 (continued)

Participant Characteristics

Characteristic	Assessment Cases		Family History Cases	
	Randomized	Cancers	Randomized	Cancers
Microcalcification	1006 (17)	282 (24)	40 (4)	1 (50)
Distortion	514 (8)	109 (9)	10 (1)	1 (50)
Asymmetric density	1837 (31)	107 (9)	26 (3)	0
None	137 (2)	7 (1)	877 (84)	0

Note.—Data are numbers of cases, and data in parentheses are percentages. Percentages were reported for known data.

* With or without ductal carcinoma in situ (DCIS).

† For invasive cancers only.

CI: 84, 90]; P = .01). In cases where the dominant radiologic feature was a mass, sensitivity of 2D mammography plus DBT (92% [95% CI: 89, 95]) was significantly increased (P = .04) compared with that of 2D mammography alone (89% [95% CI: 86, 92]). For synthetic 2D mammography plus DBT, sensitivity of 2D mammography alone (92% [95% CI: 89, 95]) was significantly higher (P = .006) than it was for synthetic 2D mammography plus DBT in cases with 11–20-mm invasive cancer. No other significant differences in sensitivity were noted.

Of the 1112 invasive cancers with reading data available, 1079 (97%) were depicted by at least one arm, and 840 (75%) were depicted by all three. There were 142 (13%) cancers missed at 2D mammography, 118 (11%) were missed at 2D mammography plus DBT, and136 (12%) were missed at synthetic 2D mammography plus DBT. For the 200 cases of ductal carcinoma in situ, the distribution was similar, with 143 (71%) cancers depicted by all three arms, eight (4%) depicted only at 2D mammography, and 14 (7%) detected at both DBT arms (but not 2D mammography).

Table 3 shows the characteristics of the cancers missed at each imaging modality and those depicted at all three modalities. The major differences are that cancers that were missed at 2D mammography tended to be 11–20 mm or have a mass as the dominant radiologic feature, whereas those that were missed at 2D mammography plus DBT were less likely to be grade 2 or occur in cases with a breast density less than 50% compared with the other two modalities.

Discussion

In this retrospective reading study, we demonstrated a clear improvement in diagnostic accuracy when DBT is used in conjunction with 2D or synthetic 2D mammography compared with 2D mammography alone, with a significant reduction (56%) in the odds of having a false-positive recall. This reduction was mainly driven by a dramatic improvement in specificity.

Subgroup analyses indicated that the sensitivity of 2D mammography plus DBT was better than that of 2D mammography for depicting grade 2 invasive cancers, 11–20-mm invasive cancers, and lesions whose dominant radiologic feature was distortion or asymmetric density. Improved depiction of spiculated masses and distortions at 2D mammography plus DBT was also reported by Rose et al (23) and attributed to clearer definition of lesion shape and margins. The addition of DBT appears to have had little impact on the depiction of cancers larger than 20 mm. Our data are consistent with those of other studies that reported that DBT is better for depicting masses rather than microcalcification as the main radiologic feature (24–26).

We observed a significant improvement in sensitivity of 2D mammography plus DBT in women aged 50–59 years and with breast density of 50% or more; however, this improved sensitivity was not seen with synthetic 2D mammography plus DBT and could reflect a shortcoming of the synthetic algorithm. It is not possible to comment on its sensitivity in women younger than 50 years old because only 29 cancers were found in women in this age group.

The significant improvement in specificity observed for 2D mammography plus DBT compared with that of 2D mammography alone is consistent with published studies (16,17,23,27). The results of the Oslo (16) and STORM (17) trials indicate the potential of 2D mammography plus DBT to reduce false-positive recall rates, although the magnitude of this reduction is likely to vary depending on the recall or arbitration policy in practice. Our study was not a screening trial, but the results are consistent with a relative improvement in specificity of 19%, suggesting that almost one in five of current false-positive recalls may be avoided by the addition of DBT. In the United States, where recall rates are often higher than those in Europe, the addition of DBT may have a greater effect. The improvement in specificity was observed irrespective of breast density, age, dominant radiologic feature, or invasive status, adding to the overall results of other studies (23,27,28).

We also compared the diagnostic performance of synthetic 2D mammography plus DBT with that of 2D mammography alone. There was no significant difference in sensitivity between synthetic 2D mammography plus DBT and 2D mammography, but specificity was significantly higher. In comparison with 2D mammography plus DBT, there was no significant difference in sensitivity or specificity. Gur et al (18) reported that synthetic 2D mammography plus DBT has lower sensitivity but comparable specificity compared with those of standard 2D mammography plus DBT; however, two recent studies that used the same software version as our study demonstrated comparable performance (29,30). Our subgroup analysis suggests that synthetic 2D mammography plus DBT and 2D mammography plus DBT had better sensitivity for depicting 11–20-mm invasive cancers than did

Table 2

Cancer Detection Rate in Three Arms of the Reading Study

	2D Mammography		2D Mammography Plus DBT		Synthetic 2D Mammography Plus DBT	
Participant Characteristic	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)	Sensitivity	Specificity
Age						
<50 years	86 (25/29)	68 (847/1241)	83 (24/29)	83 (1036/12153)	89% (25/28)	83% (953/1151)
50–59 years	87 (399/461)	54 (1646/3064)	91 (414/453)	67 (2070/3096)	88% (402/455)	69% (2034/2967)
≥60 years	88 (582/665)	57 (834/1467)	88 (583/660)	67 (980/1468)	88% (571/652)	66% (917/1400)
Breast density						
<50%	88 (715/814)	58 (2275/3930)	89 (717/806)	71 (2800/3969)	88% (706/799)	70% (2629/3747)
≥50%	86 (286/334)	57 (1002/1759)	93 (299/329)	70 (1232/1767)	87% (286/329)	72% (1220/1694)
Invasive cancer size						
1–10 mm	85 (268/315)	…	85 (262/309)	…	84% (260/308)	…
11–20 mm	86 (371/431)	…	93 (399/429)	…	92% (387/421)	…
>20 mm	93 (179/193)	…	91 (171/188)	…	90% (171/190)	…
Ductal carcinoma in situ size						
1–10 mm	80 (48/60)	…	83 (49/59)	…	85% (51/60)	…
11–20 mm	89 (41/46)	…	83 (39/47)	…	76% (35/46)	…
>20 mm	91 (86/94)	…	93 (87/94)	…	88% (82/93)	…
Cancer grade*						
1	86 (206/240)	…	89 (210/236)	…	88% (205/232)	…
2	87 (435/502)	…	91 (452/495)	…	88% (439/49)	…
3	90 (162/180)	…	88 (157/178)	…	90% (159/176)	…
Node status*						…
Negative	88 (451/513)	…	90 455/506	…	89% (448/503)	…
1–3 positive nodes	86 (248/289)	…	88 (252/287)	…	87% (250/285)	…
>3 positive nodes	86 (66/77)	…	93 (70/75)	…	88% (66/75)	…
Dominant imaging feature						
Soft-tissue mass	89 (580/650)	51 (979/1919)	92 (594/643)	67 (1287/1928)	91% (580/636)	66% (1226/1862)
Microcalcification	88 (249/282)	31 (230/745)	88 (246/279)	39 (293/750)	85% (237/278)	44% (318/723)
Distortion or ASD	71 (171/216)	64 (1363/2125)	82 (175/213)	75 (1614/2145)	82% (176/214)	76% (1540/2027)
All subjects combined	87 (1006/1155)	58 (3307/5772)	89 (1021/1142)	69 (4086/5900)	88% (998/1135)	71% (3904/5518)

Note.—Data in parentheses are the raw figures from which the percentages were calculated. ASD = asymmetrical density.

* For invasive cancers only.

Figure 2

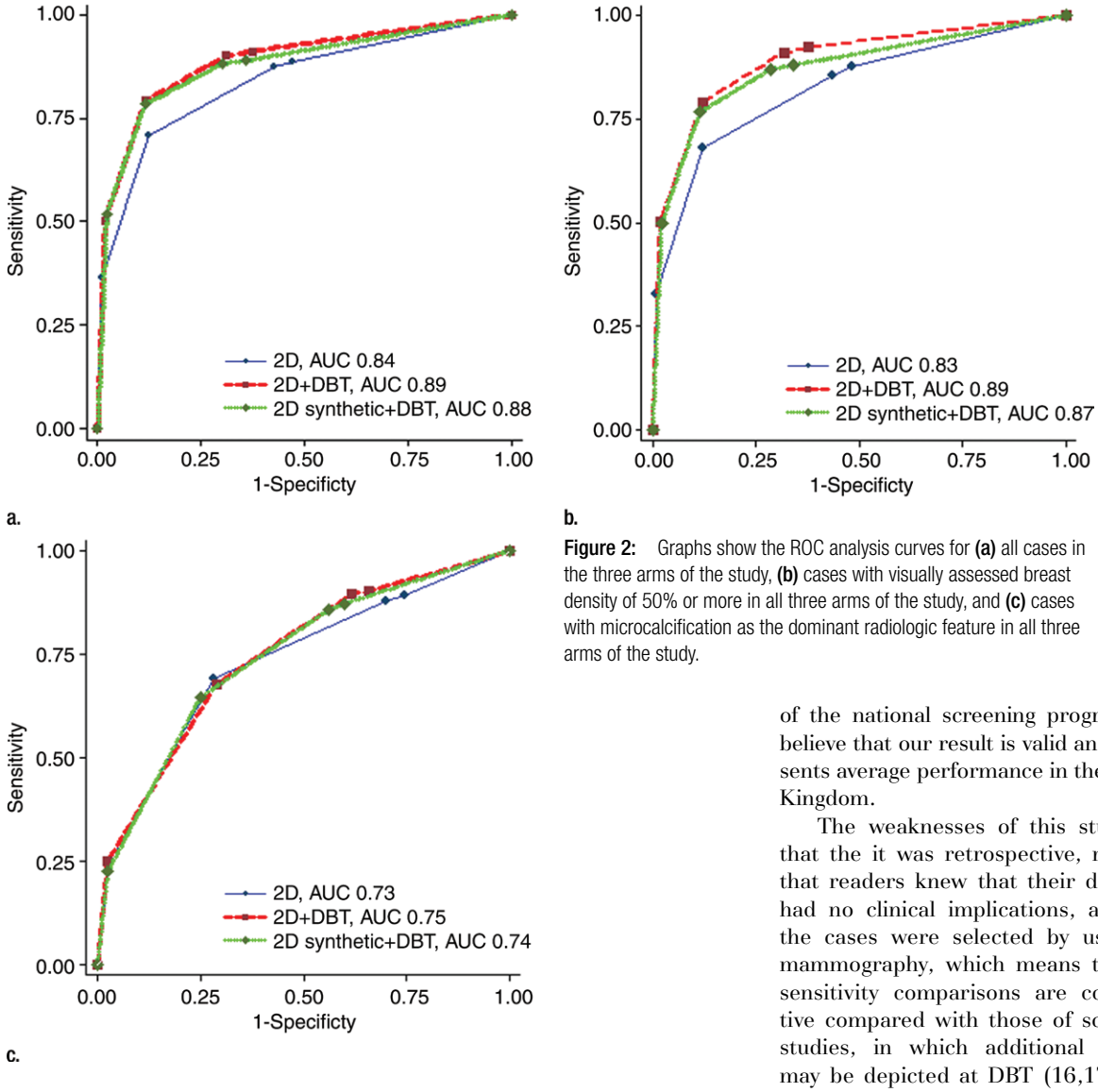

Figure 2: Graphs show the ROC analysis curves for **(a)** all cases in the three arms of the study, **(b)** cases with visually assessed breast density of 50% or more in all three arms of the study, and **(c)** cases with microcalcification as the dominant radiologic feature in all three arms of the study.

2D mammography alone. However, synthetic 2D mammography plus DBT was inferior to both 2D mammography and 2D mammography plus DBT for depicting microcalcifications and 11–20-mm ductal carcinoma in situ.

One strength of this study is that there were sufficient numbers of cancers for planned subanalyses, and the distribution of cancer cases in terms of invasion, size, grade, and lymph node status was similar to that in the United Kingdom screening program (31). Each case was independently read by three readers from three different sites, thereby avoiding carry over and reducing the risk for individual reader bias. The 26 readers were high-volume readers with varying amounts of experience. Our assessment of reader performance with the addition of DBT was conducted in a pragmatic way, and we accept that performance varies not only among readers but also for individuals from day to day. By using a large number of highly trained readers who read large volumes of screening mammograms in the regulated environment of the national screening program, we believe that our result is valid and represents average performance in the United Kingdom.

The weaknesses of this study are that the it was retrospective, meaning that readers knew that their decisions had no clinical implications, and that the cases were selected by using 2D mammography, which means that our sensitivity comparisons are conservative compared with those of screening studies, in which additional cancers may be depicted at DBT (16,17). This is also a highly enriched study, since most cases were recalled because of abnormal mammography findings.

With the use of assessment cases for this study, we effectively used 2D mammography to find all the cases in 100% of the screened women. When we used 2D mammography plus DBT in the 5% of women who were recalled from screening, the readers were, in principle, able to find all the cancers that were identified as suspicious lesions at 2D mammography for the entire population. Readers could only find cancers that were only detectable

Table 3

Characteristics of Invasive Cancers Missed at Each Imaging Modality and Found at All Modalities

	Cancer Missed at Each Modality			
Characteristic	2D Mammography	2D Mammography Plus DBT	Synthetic 2D Mammography Plus DBT	Found at All Modalities
Size				
1–10 mm	47 (39)	47 (50)	48 (47)	207 (30)
11–20 mm	60 (50)	30 (32)	34 (34)	326 (47)
>20 mm	14 (11)	17 (18)	19 (19)	155 (23)
Node status				
Negative	62 (54)	51 (56)	55 (56)	388 (60)
1–3 positive nodes	41 (36)	35 (38)	35 (35)	201 (31)
>3 positive nodes	11 (10)	5 (6)	9 (9)	55 (9)
Grade				
1	34 (29)	26 (29)	27 (27)	164 (24)
2	67 (56)	43 (48)	57 (56)	373 (55)
3	18 (15)	21 (23)	17 (17)	140 (21)
Breast density				
<50%	80 (66)	69 (73)	70 (66)	488 (70)
≥50%	42 (34)	26 (27)	36 (34)	195 (30)
Dominant imaging feature				
Mass	68 (55)	45 (47)	52 (49)	479 (69)
Microcalcifications	12 (10)	12 (13)	16 (15)	88 (13)
Distortion or ASD	42 (34)	37 (39)	37 (35)	126 (18)
None	1 (1)	1 (1)	1 (1)	4 (1)

Note.—Data are numbers of patients, and data in parentheses are percentages. ASD = asymmetrical density.

at DBT in the 5% of women who were recalled. Thus, we did not perform DBT in the other 95% of women who were not recalled, and there may be 20 times as many cancers that are only detectable at DBT in the entire screened population. In this study, the sensitivity improvement with the use of 2D mammography plus DBT was only 2%. In actual screening, we can expect up to 20 times as many cases of this type of cancer, leading to as much as a 40% increase in cancer detection for 2D mammography plus DBT compared with 2D mammography alone. This calculation shows that, when one takes account of the method of case selection, the sensitivity improvement found in this study is consistent with those of published screening studies, which reported increases of about 25%–40% (16,17). Almost all previous nonscreening studies are also affected by such case selection issues.

We found that the performance of 2D mammography plus DBT was better than that of 2D mammography in terms of specificity, with significant improvements in sensitivity for depicting specific categories of tumor, and that synthetic 2D mammography was comparable to conventional 2D mammography when used in conjunction with DBT. The improved specificity of integrated 2D mammography plus DBT was equally effective across all age groups, but the improved sensitivity seen in women aged 50–59 years and in those with dense breasts suggests that this technology could be directed toward younger women and to screening of women with a family history of breast cancer. However, the potential to reduce the burden of false-positive recalls is of considerable importance to screening programs.

Further evaluation of DBT in large-scale prospective randomized trials with subsequent round screening and collection of interval cancers is required to establish evidence-based recommendations for population screening. In particular, the diagnostic performance of synthetic 2D mammography for depicting microcalcifications could be improved. This study demonstrated that the addition of DBT is particularly valuable in certain groups of women, thereby facilitating the introduction of a more personalized approach to screening.

Disclosures of Conflicts of Interest: F.J.G. Activities related to the present article: grant from National Institute of Health Research Health Technology Assessment. Activities not related to the present article: grant from GE Healthcare and payment for lectures from Bracco. Other relationships: disclosed no relevant relationships. **L.T.** disclosed no relevant relationships. **M.G.C.G.** disclosed no relevant relationships. **P.W.** disclosed no relevant relationship. **J.C.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: developed educational presentations for Hologic. Other relationships: disclosed no relevant relationships. **K.A.D.** Activities related to the present article: grant and travel expenses from NIHR Technology Assessment Programme. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **M.J.M.** Activities related to the present article: grant from National Institute of Health Research Health Technology Assessment. Activities not related to the present article: honoraria and equipment for tomosynthesis research from Hologic. Other relationships: disclosed no relevant relationships. **H.M.D.** Activities related to the present article: disclosed no relevant relationships. Activities not related to the present article: travel expenses for trial study meeting. Other relationships: disclosed no relevant relationships. **Y.Y.L.** Actvities related to the present article: grant from National Institute of Health Research Health Technology Assessment. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **T.S.** disclosed no relevant relationships. **S.M.A.** disclosed no relevant relationships. **O.M.** disclosed no relevant relationships. **K.C.Y.** Activities related to the present article: grant from National Institute of Health Research Health Technology Assessment. Activities not related to the present article: disclosed no relevant relationships. Other relationships: disclosed no relevant relationships. **S.W.D.** disclosed no relevant relationships.

Disclaimers: F.J.G. had full access to all the data in the study and final responsibility for the decision to publish. The sponsor had no role in the study design, data collection, analysis and interpretation, or writing of the report.

References

1. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380(9855):1778–1786.
2. Michell MJ. Breast screening review: a radiologist's perspective. Br J Radiol 2012;85(1015):845–847.
3. Duncan KA, Needham G, Gilbert FJ, Deans HE. Incident round cancers: what lessons can we learn? Clin Radiol 1998;53(1):29–32.
4. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003;138(3):168–175.
5. 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.
6. Chiu SY, Duffy S, Yen AM, Tabár L, Smith RA, Chen HH. Effect of baseline breast density on breast cancer incidence, stage, mortality, and screening parameters: 25-year follow-up of a Swedish mammographic screening. Cancer Epidemiol Biomarkers Prev 2010;19(5):1219–1228.
7. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. Radiology 1997;205(2):399–406.
8. Dobbins JT 3rd. Tomosynthesis imaging: at a translational crossroads. Med Phys 2009;36(6):1956–1967.
9. Sechopoulos I. A review of breast tomosynthesis: part I—the image acquisition process. Med Phys 2013;40(1):014301.
10. Sechopoulos I. A review of breast tomosynthesis: part II—image reconstruction, processing and analysis, and advanced applications. Med Phys 2013;40(1):014302.
11. Diekmann F, Bick U. Breast tomosynthesis. Semin Ultrasound CT MR 2011;32(4):281–287.
12. Helvie MA. Digital mammography imaging: breast tomosynthesis and advanced applications. Radiol Clin North Am 2010;48(5):917–929.
13. Baker JA, Lo JY. Breast tomosynthesis: state-of-the-art and review of the literature. Acad Radiol 2011;18(10):1298–1310.
14. Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. Breast 2013;22(2):101–108.
15. Alakhras M, Bourne R, Rickard M, Ng KH, Pietrzyk M, Brennan PC. Digital tomosynthesis: a new future for breast imaging? Clin Radiol 2013;68(5):e225–e236.
16. 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.
17. 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.
18. Gur D, Zuley ML, Anello MI, et al. Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. Acad Radiol 2012;19(2):166–171.
19. 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.
20. McNEMAR Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12(2):153–157.
21. Breslow NE, Day NE. Statistical methods in cancer research: vol 1—the analysis of case-control studies. Lyon, France: IARC, 19812.
22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44(3):837–845.
23. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. AJR Am J Roentgenol 2013;200(6):1401–1408.
24. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. AJR Am J Roentgenol 2007;189(3):616–623.
25. Spangler ML, Zuley ML, Sumkin JH, et al. Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: a comparison. AJR Am J Roentgenol 2011;196(2):320–324.
26. Rafferty EA, Park JM, Philpotts LE, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. Radiology 2013;266(1):104–113.
27. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology 2013;269(3):694–700.
28. Bernardi D, Ciatto S, Pellegrini M, et al. Prospective study of breast tomosynthesis as a triage to assessment in screening. Breast Cancer Res Treat 2012;133(1):267–271.
29. Zuley ML, Guo B, Catullo VJ, et al. Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. Radiology 2014;271(3):664–671.
30. 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.
31. NHS Breast Screening Programme 2012 Annual Review. http://www.cancerscreening.nhs.uk/breastscreen/publications/2012review.html. Published 2012. Accessed July 10, 2014.