A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis

M.J. Michell a,*, A. Iqbal a, R.K. Wasan a, D.R. Evans a, C. Peacock a, C.P. Lawinski b, A. Douiri c, f, R. Wilson d, P. Whelehane e

a Breast Radiology, King’s College Hospital, London, UK
b KCARE, King’s College Hospital, London, UK
c Department of Primary Care and Public Health Sciences, King’s College London, London, UK
d Department of Clinical Radiology, The Royal Marsden NHS Foundation Trust, Surrey, UK
e Medical Research Institute, University of Dundee, Ninewells Hospital & Medical School, Dundee, UK
f National Institute for Health Research Comprehensive Biomedical Research Centre, Guy’s and St. Thomas’ NHS Foundation Trust, London, UK

AIM: To measure the change in diagnostic accuracy of conventional film-screen mammography and full-field digital mammography (FFDM) with the addition of digital breast tomosynthesis (DBT) in women recalled for assessment following routine screening.

MATERIALS AND METHODS: Ethics approval for the study was granted. Women recalled for assessment following routine screening with screen-film mammography were invited to participate. Participants underwent bilateral, two-view FFDM and two-view DBT. Readers scored each lesion separately for probability of malignancy on screen-film mammography, FFDM, and then DBT. The scores were compared with the presence or absence of malignancy based on the final histopathology outcome.

RESULTS: Seven hundred and thirty-eight women participated (93.2% recruitment rate). Following assessment 204 (26.8%) were diagnosed as malignant (147 invasive and 57 in-situ tumours), 286 (37.68%) as benign, and 269 (35.4%) as normal. The diagnostic accuracy was evaluated by using receiving operating characteristic (ROC) and measurement of area under the curve (AUC). The AUC values demonstrated a significant (p = 0.0001) improvement in the diagnostic accuracy with the addition of DBT combined with FFDM and film-screen mammography (AUC = 0.9671) when compared to FFDM plus film-screen mammography (AUC = 0.8949) and film-screen mammography alone (AUC = 0.7882). The effect was significantly greater for soft-tissue lesions [AUC was 0.9905 with the addition of DBT and AUC was 0.9201 for FFDM with film-screen mammography combined (p = 0.0001)] compared to microcalcification [with the addition of DBT (AUC = 0.7920) and for FFDM with film-screen mammography combined (AUC = 0.7843; p = 0.3182)].

CONCLUSION: The addition of DBT increases the accuracy of mammography compared to FFDM and film-screen mammography combined and film-screen mammography alone in the assessment of screen-detected soft-tissue mammographic abnormalities.

© 2012 The Royal College of Radiologists. Published by Elsevier Ltd. All rights reserved.
Introduction

X-ray mammography using analogue film-screen and full-field digital mammography (FFDM) has been shown to be effective for both routine screening and symptomatic breast diagnosis. The accuracy of conventional mammography is limited by anatomical noise resulting from the superimposition of normal structures. In clinical practice, this may affect both sensitivity and specificity — cancer detection may be limited, particularly in younger women and those with dense breast patterns because mammographic evidence of the tumour may be completely or partially obscured. Conversely, overlying normal structures may produce an appearance with conventional two-dimensional mammography, which is suspicious for cancer and leads to recall for further assessment following routine screening. Digital breast tomosynthesis (DBT) has the potential to improve the accuracy of mammography by enabling the reader to view x-ray images of the breast tissue as a series of thin reconstructed sections, so overcoming the problem of overlying tissues on conventional two-dimensional images.

The technical features of DBT have been described in detail in previously published work. Recent articles and presentations at scientific meetings have addressed the application of DBT in clinical practice. Some data have shown an improvement in specificity and Breast Imaging Reporting and Data System (BI-RADS) classification of lesions using DBT alone or in combination with FFDM. However, there remains some uncertainty about whether and how DBT should be used in routine screening, in the assessment and diagnostic work-up of mammographically detected abnormalities found in routine screening and in women presenting with breast symptoms.

The present study examines the impact of the addition of FFDM and DBT on the diagnostic accuracy of mammography in a population of women recalled following film-screen mammography alone.

Materials and methods

The aim of the study was to assess the potential additional diagnostic accuracy over film-screen mammography. FFDM and DBT in a population of women recalled for assessment in a screening programme. The study primarily measures the specificity of these two additional digital imaging techniques in an unblinded clinical setting. This design was required as at the time DBT was not approved for routine use in breast screening in the UK.

Research ethics committee approval for the study was obtained in November 2008 and women were invited to participate from January 2009 to July 2010. Women eligible for recruitment were those recalled for further assessment of a mammographic abnormality found on routine film-screen mammographic screening. The screening examinations were carried out either on a mobile or static screening site. All screening mammograms were double read by radiologist or radiographer film readers meeting the UK National Health Service Breast Screening Programme (NHBSBSP) standard of reading more than 5000 mammograms per annum. In cases of discordance between reader one and reader two, the decision to recall or not was made following discussion and consensus by two or more readers. The diagnostic work-up at assessment was carried out by one of the team of five specialist breast radiologists, according to NHBSBSP screening assessment guidelines.

Women eligible for the study were provided with written information at the time of invitation to attend the breast assessment clinic. On attendance at the clinic, the study was discussed with eligible women and written consent was obtained from those who agreed to participate.

Participating women underwent bilateral two view medial–lateral–oblique (MLO) and cranio-caudal (CC) FFDM and DBT within a single compression episode for each projection. The mammography machine was a FFDM unit with tomosynthesis capability (Hologic Selenia Dimensions, Hologic, Bedford, MA). Eleven low-dose projection images are acquired during movement of the x-ray tube through a 15° arc. Reconstructed 1 mm sections in a plane parallel to the digital image receptor surface are then produced for viewing.

During the period of this study routine quality control measurements were made on the Dimensions system following the methods and protocols suggested by the Institute of Physics and Engineering in Medicine (IPEM) and the NHBSBSP. For FFDM the mean glandular dose (MGD) to a standard breast was calculated using software provided by the National Centre for the Coordination of Physics in Mammography. The dose levels varied between 1.37 and 1.57 mGy. The range was due to small differences in machine performance over time, and a series of software upgrades provided by the manufacturer, which resulted in small changes in dose.

Currently, a national protocol for the measurement of breast dose for DBT has not been finalized. Therefore, DBT dose levels were estimated using data provided by the manufacturer (Hologic) and varied between 1.66 and 1.90 mGy for a standard breast.

The screen-film, FFDM, and DBT examinations were viewed sequentially by the radiologist carrying out the assessment according to a standard protocol and data were recorded on the research data collection form. The film-screen images were viewed first, followed by the FFDM and then the DBT images. The radiologist was able to use the tools available on the digital mammography workstation for adjustment of brightness and contrast, magnification and black/white image inversion for the FFDM and DBT images. DBT images were viewed as 1 mm reconstructed sections and the facility to increase the section thickness was available.

The following data were recorded for each case: mammography features of any lesion resulting in recall to assessment including mammographic sign (circumscribed lesion, spiculate lesion, microcalcification, parenchymal distortion, asymmetry), size in millimetres, suspicion level according to the Royal College of Radiologists Breast Group...
mammography classification system (1 = normal, 2 = benign, 3 = probably benign, 4 = suspicious, 5 = malignant), site on both the MLQ and CC views and mammographic breast density (fatty, glandular, dense).\textsuperscript{16}

The final outcome was recorded as normal/benign, malignant invasive or malignant non-invasive. Cases recorded as normal/benign included all cases where further imaging and clinical assessment showed no features to suggest malignancy and those cases that underwent needle biopsy or surgery showing benign findings. Malignant invasive and malignant non-invasive results were based on final surgical histology. The diagnosis and management of all cases undergoing needle biopsy and/or surgery were discussed at the weekly prospective multidisciplinary meetings.

The inferential statistical analysis was performed using receiver operative characteristic (ROC) methods to evaluate the results. Sub-analysis was carried out for mammographic feature and parenchymal density. ROC analysis was performed to assess and compare the diagnostic accuracy of the three imaging methods by calculating the area under the curve (AUC) as described by Hanley et al.\textsuperscript{17,18} A ROC graph, with the y-axis representing the true-positive rate and the x-axis representing false-negative rate, was used to plot the individual and summary points of sensitivity and specificity. All statistical analyses were carried out with Stata Software version 11.0.\textsuperscript{19}

\section*{Results}

Seven hundred and ninety-two women recalled to assessment because of a mammographic abnormality following routine screening using film-screen mammography were invited to participate in the study over an accrual period of 18 months. Women recalled because of a clinical symptom in whom the screening mammograms were normal were not invited to participate. Seven hundred and thirty-eight (of the 792 (93.2%) women agreed to participate in the study.

Seven hundred and ninety-two lesions were identified in 738 participants. In 17 patients, there were two lesions (two patients with bilateral benign lesions, two patients with bilateral malignant lesions, and one patient with ipsilateral benign and malignant lesions. The remaining 12 patients had two ipsilateral malignant lesions.) In two patients, there were three lesions (one patient with two cancers in the contralateral breast, one patient with three cancers in one breast). Two hundred and four of 759 (26.8%) lesions were malignant: 147/204 (72%) were invasive cancer, 57/204 (27.9%) were ductal carcinoma in situ (DCIS). Following assessment, 555/759 (73%) lesions were found to be either normal or benign — these included cysts and lesions with a benign outcome on either needle biopsy or surgical histology.

The additional information provided by DBT significantly improved the accuracy compared to film-screen mammography alone and FFDM with film-screen mammography combined. The diagnostic accuracy of the imaging methods as reflected by the mean (±SE) of AUC demonstrated that for film-screen mammography alone, the AUC was $0.7882 \pm 0.0198$, 95% CI $0.74945–0.82702$; for FFDM with film-screen mammography combined the AUC was $0.8949 \pm 0.0124$, 95% CI $0.87061–0.91915$; and with the addition of DBT the AUC was $0.9671 \pm 0.0050$, 95% CI $0.95732–0.97683$. The difference in AUC between DBT and FFDM with film-screen mammography combined was $0.0722 (p = 0.0001)$, and the difference in AUC between DBT and film-screen mammography alone was $0.1789 (p = 0.0001); \text{Fig 1}$. The sensitivity (absolute and complete), specificity, positive predictive value (PPV), and negative predictive value (NPV) for FFDM and DBT are shown in the Table 1.

The distribution of malignant ($n = 204$) and benign ($n = 555$) lesions according to mammography score and imaging method is shown in Figs 2 and 3. Of the 204 cancers, 70/204 (34.3%) were classified as malignant (M5) on film-screen mammography; this improved to 81/204 (39.7%) on FFDM and 119/204 (58.3%) with the addition of FFDM and DBT, respectively. The addition of DBT resulted in fewer malignant lesions being classified as M3/M4 compared to film-screen mammography and FFDM combined. Six of the 204 (2.9%) cancers were only visualized on DBT. Two lesions were ipsilateral additional cancers, two lesions were contralateral cancers and one lesion found in the ipsilateral breast in a patient recalled for benign microcalcification proven by needle biopsy.

Women with benign findings were discharged from the assessment clinic to continue routine screening. The follow-up period was 18 to 36 months. In one case, a 10 mm invasive grade 1 contralateral cancer was diagnosed and treated 24 months following initial diagnosis and treatment of the screen-detected cancer. This was detected during a routine retrospective audit of screen-detected cancers and was not diagnosed on routine follow-up on FFDM. No other cases are known to have presented with an interval breast cancer through routine interval cancer data collection by the service.

The addition of DBT resulted in more normal/benign cases being classified as (M1/M2) [412/555 (74.2%)]
The distribution of malignant lesions (Figure 2) compared to film-screen mammography alone [21/555 (3.78%)] and FFDM with film-screen mammography combined [283/555 (50.9%)]. Fig 6 shows the distribution by mammography score on DBT of malignant and benign soft-tissue lesions classified as M3 on FFDM. There were no malignant cases classified as M1 or M2 on DBT. Twenty-two of 218 (10.1%) malignant lesions were classified as either suspicious or malignant (M4/M5) counted as positive.

Sub-analysis by mammographic density did not show any significant difference for the detection of cancers with radiographically fatty, dense, or glandular breasts (Table 2). Analysis according to mammographic features showed that in soft-tissue lesions there was a significant difference in accuracy between FFDM with film-screen mammography combined and the addition of DBT (Fig 4); AUC 0.9201 ± 0.0133 (95% CI 0.8941–0.94614) and 0.9905 ± 0.0024 (95% CI 0.98585–0.99519), respectively (difference in AUC 0.0704; \( p = 0.0001 \)). However, for microcalcification, there was no significant difference between the two imaging techniques (Fig 5). Area under the ROC curve for DBT was 0.7920 ± 0.0320 (95% CI 0.72930–0.85460) and for FFDM with film-screen mammography combined was 0.7843 ± 0.0322 (95% CI 0.72115–0.84752; difference 0.0077; \( p = 0.3182 \)).

### Discussion

Previously published studies have demonstrated that the use of DBT in addition to FFDM may improve the accuracy in the classification of mammographic features, as measured using BIRADS or similar mammography classification systems. The improvements in diagnostic accuracy are due to the reduction in anatomical noise of the image and greater ability of the radiologist to perceive, view, and interpret the details of both normal and abnormal features on x-ray mammograms.

In this study of women recalled for further assessment following routine screening, FFDM was more accurate than screen-film mammography. This is consistent with the findings of trials comparing the accuracy of film-screen mammography and FFDM. The addition of DBT, however, showed a significant improvement in accuracy compared to both film-screen mammography alone and film-screen mammography and FFDM combined. The improvement in visualization of subtle signs using DBT enabled the reader to classify indeterminate lesions found on FFDM more accurately as either more suspicious for malignancy or more likely to be benign. In this study, additional lesions were detected on DBT that had not been visualized with film-screen mammography with FFDM combined. Six additional cancers, all spiculate lesions, were found in four patients. In all of these cases, the additional foci were identified on second-look ultrasound, aided by the three-dimensional information provided by DBT. In 126/218 (57.8%) cases with low-suspicion (M3) lesions on film-screen mammography with FFDM combined, DBT correctly demonstrated that the case was normal or benign. The normal breast parenchyma has a different appearance, enabling the reader to classify indeterminate lesions found on FFDM more accurately as either more suspicious for malignancy or more likely to be benign.

The improvement in diagnostic accuracy with the additional information from DBT was limited to soft-tissue lesions, as might be expected from the reduction in the effect of anatomical noise using DBT. For microcalcification, DBT did not improve accuracy. DBT was not compared with fine-focus magnification mammography in this study. The change in accuracy was examined prospectively. In order to compare the different techniques directly, the DBT images need to be read independently of the FFDM images, and the present authors are currently engaged in a multi-reader blinded comparison study.

There are significant implications for both screening and diagnostic practice from this and other studies. DBT may increase the efficiency and effectiveness of screening through improvement in both specificity and sensitivity. Better visualization of normal and benign features may improve the ability of the film reader to confidently diagnose cases without malignancy and thereby reduce the recall rate, with savings in resources for the healthcare system and fewer women suffering the anxiety and inconvenience associated with recall. Conversely, improved visualization of abnormal features may improve sensitivity for the detection of small cancers with subtle signs and decrease the false-negative interval cancer rate. However, these potential improvements in specificity and sensitivity will need to be carefully examined in large-scale trials. There are increased costs associated with DBT, related to both equipment and reading time, but a cost-effectiveness

### Table 1

Computed values for absolute sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of histology-proven malignant lesions.

<table>
<thead>
<tr>
<th></th>
<th>Full-field digital mammography (%)</th>
<th>Digital breast tomosynthesis (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute sensitivity</td>
<td>39.7</td>
<td>58.3</td>
</tr>
<tr>
<td>Complete sensitivity</td>
<td>97.5</td>
<td>100</td>
</tr>
<tr>
<td>Specificity</td>
<td>51</td>
<td>74.2</td>
</tr>
<tr>
<td>PPV</td>
<td>42.3</td>
<td>58.8</td>
</tr>
<tr>
<td>NPV</td>
<td>98.3</td>
<td>100</td>
</tr>
</tbody>
</table>

\( a \) Absolute sensitivity = M5 counted as positive.

\( b \) Complete sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) = M3 + M4 + M5 counted as positive.

Figure 2 The distribution of malignant lesions (n = 204) according to mammography score for film-screen, FFDM, and DBT.
analysis was beyond the scope of this study and should be addressed by future research. The equipment from a single manufacturer was used in conducting this trial. The technique and features of DBT machines vary considerably between different manufacturers and clinical performance results are, therefore, not necessarily generalizable.

In the diagnostic and assessment setting, imaging work-up has traditionally involved both ultrasound and supplementary mammography, including spot compression and magnification views. DBT provides detailed mammographic information for soft-tissue lesions, demonstrates the whole breast, and is not dependent on the accurate positioning over the target lesion required for spot compression views. Law\textsuperscript{20} investigated the breast dose from magnification films on a range of x-ray units of various designs with a nominal magnification factor of 1.8. The ratio derived between the dose level using magnification geometry and conventional contact geometry was 2.2, giving an average magnification film dose of 5 mGy (based on an average screen film MGD of 2.3 mGy). Thus the use of FFDM with DBT can provide a considerable dose advantage over the use of screen-film magnification techniques.

A comparison of DBT with ultrasound was beyond the scope of this study. Further studies are required to assess the added value of DBT in the diagnostic work-up and local staging of malignancy in the context of the full range of current techniques, including ultrasound, magnification and spot compression mammography, and MRI.

The conclusions that can be reached from this prospective preliminary study are limited by the method; the readers were unblinded, they read and scored the DBT

![Figure 3](image)

**Figure 3** The distribution of the benign and normal lesions (n = 555) according to mammography score for film-screen, FFDM, and DBT.

<table>
<thead>
<tr>
<th></th>
<th>Fatty breast</th>
<th></th>
<th>Dense and glandular breasts</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FFDM</td>
<td>DBT</td>
<td>FFDM</td>
<td>DBT</td>
</tr>
<tr>
<td>AUC</td>
<td>0.9342</td>
<td>0.9903</td>
<td>0.8863</td>
<td>0.9624</td>
</tr>
<tr>
<td>SE</td>
<td>0.0202</td>
<td>0.0062</td>
<td>0.0143</td>
<td>0.0059</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.894 – 0.973</td>
<td>0.978 – 1.000</td>
<td>0.858 – 0.914</td>
<td>0.950 – 0.973</td>
</tr>
<tr>
<td>p-Values</td>
<td>0.0002</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FFDM, full-field digital mammography; DBT, digital breast tomosynthesis; AUC, area under the curve.

![Figure 4](image)

**Figure 4** ROC curve analysis for soft-tissue lesions (n = 603).

![Figure 5](image)

**Figure 5** ROC curve analysis for micro-calciﬁcation (n = 156).
images with knowledge of the film-screen mammography and FFDM findings. This study has, therefore, demonstrated improved diagnostic accuracy of film-screen mammography and FFDM combined compared to film-screen mammography alone, and improved accuracy when DBT is added to film-screen mammography and FFDM, in the interpretation of soft-tissue mammographic abnormalities in women recalled following routine screening.

Acknowledgement

This project was supported with funding from the NHSBSP.

References


15. National Centre for the Coordination of Physics in Mammography (NCCPM).


