Digital Breast Tomosynthesis

Whelehan, Patsy; Ali, Kulsam; Vinnicombe, Sarah; Ball, Graham; Cox, Julie; Farry, Paul

Published in:
British Journal of Radiology

DOI:
10.1259/bjr.20201105

Publication date:
2021

Document Version
Peer reviewed version

Digital Breast Tomosynthesis: sensitivity for cancer in younger symptomatic women

(Shortened title: DBT sensitivity in younger symptomatic women)

Full paper

Patsy Whelehan, MSc[1, 2], Kulsam Ali, PhD[1], Sarah Vinnicombe, FRCR[3], Graham Ball, PhD[4], Julie Cox, RCR[5], Paul Farry, FRCR[6], Maggie Jenkin, MSc[7], Keith Lowry, FRCR[8], Stuart A McIntosh, FRCS[9], Rachel Nutt, PhD[1], Rachel Oeppen, FRCR[10], Michael Reilly, FRCR[6], Michaela Stahnke, FRCR[10] Jim Steel, FRCR[7], Yee Ting Sim, FRCR[2], Violet Warwick, PhD[1], Louise Wilkinson, FRCR[11], Dimitrios Zafeiris, PhD[4], Andy Evans, FRCR[1]

1. School of Medicine, University of Dundee, Mailbox 4, Ninewells Hospital & Medical School, Dundee, DD1 9SY, UK; p.j.whelehan@dundee.ac.uk **(Corresponding author)**
2. NHS Tayside, Dundee, UK
3. Gloucestershire Hospitals NHS Foundation Trust, Cheltenham, UK
4. Nottingham Trent University, Nottingham, UK & Intelligent OMICS Ltd, Nottingham, UK
5. South Tyneside and Sunderland NHS Foundation Trust, Sunderland, UK
6. Western Health and Social Care Trust, Londonderry, UK
7. Derriford Hospital, Plymouth, UK
8. Belfast Health and Social Care Trust, Belfast, UK
9. Queen's University, Belfast, UK
10. University Hospital Southampton NHS Foundation Trust, Southampton, UK
11. Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Declarations of interest

Eighteen of 19 co-authors have no conflict of interest to declare. One author (SV) provides paid expert feedback to Siemens Healthcare GmbH on product design.

Funding

The equipment upgrade required for the study was funded, for the lead site only, by Ninewells Cancer Campaign. Further small grants to support the running of the study were provided by Tenovus, the Sir John Fisher Foundation, and a local fundraiser for the lead site, Mrs Fiona Edwards. None of the funders had input to the design or reporting of the study.

The study was supported in-kind by the University of Dundee, by the host NHS organisations, and by the NHS clinical research support services and networks in England, Scotland and Northern Ireland.

Acknowledgments

We gratefully acknowledge the patients who participated in this study.

A large number of administrative, clinical and trials staff at the participating sites supported the study; their invaluable contributions are gratefully acknowledged.

We thank Dr Melissa Hill of Volpara Health Technologies Ltd. for helpful comments on the draft manuscript, and Dr Mark Halling-Brown and colleagues at the National Co-ordinating Centre for Physics in Mammography for providing and tailoring the MedXViewer software and for associated data management support.

Abstract

Objectives

Full-field digital mammography (FFDM) has limited sensitivity for cancer in younger women with denser breasts. Digital breast tomosynthesis (DBT) can reduce the risk of cancer being obscured by overlying tissue.

The primary study aim was to compare the sensitivity of FFDM, DBT and FFDM-plus-DBT in women under 60 years old with clinical suspicion of breast cancer.

Methods

This multicentre study recruited 446 patients from UK breast clinics. Participants underwent both standard FFDM and DBT. A blinded retrospective multi-reader study involving twelve readers and 300 mammograms (152 malignant and 148 benign cases) was conducted.

Results

Sensitivity for cancer was 86.6% with FFDM (95% CI: 85.2-88.0%), 89.1% with DBT (95% CI: 88.2-90%), and 91.7% with FFDM+DBT (95% CI: 90.7-92.6%). In the densest breasts, the maximum sensitivity increment with FFDM+DBT over FFDM alone was 10.3%, varying by density measurement method. Overall specificity was 81.4% with FFDM (95% CI: 80.5-82.3%), 84.6% with DBT (95% CI: 83.9-85.3%), and 79.6% with FFDM+DBT (95% CI: 79.0-80.2%). No differences were detected in accuracy of tumour measurement in unifocal cases.

Conclusions

Where available, DBT merits first-line use in the under 60 age group in symptomatic breast clinics, particularly in women known to have very dense breasts.

Advances in knowledge

This study is one of very few to address the accuracy of digital breast tomosynthesis in symptomatic rather than screening patients. It quantifies the diagnostic gains of DBT in direct comparison with

standard digital mammography, supporting informed decisions on appropriate use of DBT in this population.

Introduction

Full-field digital mammography (FFDM) has limited sensitivity for breast cancer in younger women with denser breasts. Subgroup analysis in the DMIST (Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening) trial showed that the sensitivity of FFDM in women under 50 with dense breasts was only 59%[1]. Because DMIST was a screening trial, the cancers would have been smaller than those found in a symptomatic population. Lower mammographic sensitivity has been demonstrated in younger women presenting symptomatically in earlier studies with film-screen mammography: 67% on average in women under 60 years versus 87% in those aged 60-70[2]. Although evidence on sensitivity rates of FFDM in symptomatic populations is limited, a study in Germany has demonstrated that young age and dense breasts remain risk factors for false negative mammography in symptomatic women in the digital era[3].

Digital breast tomosynthesis (DBT) has the potential to alleviate the problem of cancers being masked on FFDM by the dense breast tissue which is characteristic of younger breasts, because the technology partially separates overlapping structures[4]. It has been shown in a sample of patients with dense breasts and either screen-detected or symptomatically presenting lesions that DBT has a sensitivity of about 88% - a 10% increment over FFDM[5]. Sensitivity and other diagnostic performance parameters have rarely been compared in exclusively symptomatic patient samples. Two such studies have now been published but both involved only the Hologic Selenia Dimensions equipment (Hologic Inc., Marlborough, MA, USA)[6,7]. The study by Bian and colleagues, in women with dense breasts, found that sensitivity increased from 58.8% with FFDM to 68.1% with DBT, although no statistical test of this difference is reported[6]. In their sample of symptomatic patients not selected by breast density, Tang and colleagues found statistically significant improvements in sensitivity with FFDM plus DBT compared to FFDM alone, which they reported separately for each of two radiologists[7]. The sensitivity increments were in the order of 20%, with little change in specificity.

Because DBT technology differs significantly between vendors, results from single-vendor studies, such as these two, are not necessarily generalizable to other equipment.

The aim of our multicentre study was primarily to compare the sensitivity for breast cancer of DBT, FFDM, and the two combined, using the Siemens Mammomat Inspiration unit (Siemens Healthcare GmbH, Erlangen, Germany) in women aged under 60 years presenting with symptoms or signs of possible breast cancer. Secondary aims were to compare specificity, differential sensitivity according to mammographic breast density and breast cancer type, and to compare accuracy for assessing tumour size.

Materials and methods

Approvals

The study was approved by the National Health Service (NHS) Research Ethics Service and received management approval in all participating institutions. The study was registered on a public database [details redacted for blinded review].

Patients

Patients were recruited from specialist breast multidisciplinary clinics in five UK hospitals, to which they had been referred for investigation of breast symptoms. They were eligible if female, aged under 60 years, if they had an abnormality which the clinician performing physical examination graded as having a greater than 20% likelihood of malignancy, and if they were referred for and agreed to mammography. Patients classified as normal or benign on clinical examination ("P"-score 1 or 2 on a scale of 1-5) were excluded. The purpose was to achieve the requirements of the power calculation to detect a difference in sensitivity, while avoiding excessive recruitment overall. The upper age limit was informed by previous research on the sensitivity of mammography in symptomatic women of different age groups[2]. Patients aged over 25 but below the local age threshold for mammography to be used as a first-line imaging procedure (usually 40 years) but in

whom ultrasound examination gave sufficient cause for suspicion to justify mammography were also eligible, irrespective of clinical suspicion.

Patients were excluded if they lacked capacity to give informed consent, were pregnant or lactating, or if they had obvious locally advanced breast cancer or severe co-morbidities expected to preclude surgical treatment. During the recruitment period, the DBT function was not cleared by the manufacture for use on patients with breast implants. In some clinics, eligible patients were not approached because of logistical issues, e.g. equipment breakdown, no radiologist with DBT reporting training available in the clinic, or no-one available to take written informed consent. It was not feasible to keep records of patients who met the eligibility criteria but were not approached.

Following written informed consent, all participants underwent a combined examination consisting of bilateral FFDM and DBT on a Siemens Mammomat Inspiration unit. Both standard care imaging and DBT findings were taken into account in the real-time diagnostic triple assessment process.

Sample size

A power calculation for a chi-squared variance test was performed using Statistica version 8 (StatSoft Inc., Tulsa, OK, USA), assuming a population variance of 0.2. To detect a 25% reduction in the FFDM occult rate, from an expected 20% for mammography to 15% using DBT or FFDM+DBT, with a statistical power of 0.8, it was calculated that 150 participants with cancer were required. The numbers used reflected an element of uncertainty regarding the variance of the sample. The size of the difference to be detected was chosen based on the chief investigator's professional judgement on the level of benefit required to influence clinical practice. In order to include the required number of participants with cancer, 446 participants were recruited in total, of whom 154 had cancer.

Retrospective multi-reader study

The retrospective reading exercise which is the subject of this manuscript included all the recruited cancer cases except for two in which we could not retrieve the full imaging dataset from the

recruiting site (n=152). Randomised selection of normal and benign cases was undertaken to provide a total of 300 cases for inclusion in the reader study. Further details of the sample are provided in Figure 1 and in the Results section text. Randomised assignment of the 300 cases into batches of 50 was undertaken, which resulted in similar distributions per batch of patient age, and cancer, benign and normal cases.

The FFDM-only, DBT-only, and FFDM+DBT images for each batch of 50 cases were separately packaged with viewing software, and each batch of 50 cases was assigned to two readers from a pool of twelve. Thus each case was read twice under each of the three conditions (300 cases x 3 conditions x 2 readers = 1,800 exam-reads in total). No reader read the same case twice. All readers read a total of three batches, one each of FFDM-only, DBT-only and FFDM+DBT. Allocation of specific batches to readers was randomised, as was the order in which they read their FFDM, DBT and FFDM+DBT batches.

All readers were trained and clinically experienced with Siemens DBT. Eleven were consultant radiologists and one was a radiographer. Radiographers in the UK are able to undertake mammography interpretation and reporting, subject to recognised additional training and terms of employment.

Data collection in the reader study was facilitated by a computer-based tool designed for observer studies in mammography and tomosynthesis (Medical Extensible Viewer - “MedXViewer” - National Co-ordinating Centre for Physics in Mammography, Guildford, UK). MedXViewer enabled display of the images on 5 mega-pixel mammography reporting monitors with on-screen tools for the readers to mark regions of interest (ROI), measure lesions where applicable, and describe abnormalities. In line with real-life practice, readers were provided with information on clinical presentation and patient age. They were instructed to ignore inconsequential benign radiological features that they would pass over in the clinic. For significant lesions, readers recorded a suspicion score according to

the UK 1-5 scale, where 1 is normal and 5 is malignant[8]. Scores of 3-5 were considered malignant in the analysis of sensitivity for cancer (sometimes known as "complete" sensitivity).

In the FFDM reading condition, we also asked readers to provide a BI-RADS® (5th edition) breast density score: (a: The breasts are almost entirely fatty; b: There are scattered areas of fibroglandular density; c: The breasts are heterogeneously dense, which may obscure small masses; d: The breasts are extremely dense, which lowers the sensitivity of mammography[9]) and to use an on-screen 0-100 mm visual analogue scale (VAS) to assign an area-based percentage mammographic density to the mammogram, based on their impression of all images in the examination. The FFDM images were also subjected to software assessment of percent volumetric breast density using Volpara® Data Manager™ software (Volpara Solutions Ltd., Wellington, New Zealand), algorithm version 1.5.0. The value used for analysis was the mean of the per-image output values for the images in the FFDM examination of the non-cancer-bearing breast. Bilateral cancer cases (n=5) and participants with cancer with only one breast examined (n=2) were therefore excluded from this sub-analysis. Volpara data were missing for six cases because the raw DICOM images required for software processing were unobtainable.

Readers measured lesion size using an on-screen ruler. For analysis of the relative accuracy of malignant lesion measurements in patients with unifocal cancer, only the FFDM alone and DBT alone reading conditions were included. Reader measurements were compared to the histopathological whole tumour diameter (WTD). Patients treated with neoadjuvant systemic therapy were excluded from the disease-extent analyses.

Ground Truth

The ground truth was established from the results of triple assessment (clinical examination, medical imaging and histopathological examination as applicable). Using the MedXViewer software, the mammograms for each case were annotated and the ground truth recorded (malignant, benign or

normal) by one of two senior consultant radiologists from the pool of readers. They were provided with both the FFDM and DBT images and the triple assessment information to enable them to identify and classify the lesions. They marked each lesion by a generously-sized freehand ROI on each view where it was visible, on the two modalities. If a malignant lesion known to be present was occult on FFDM and DBT, they marked its location based on the information available from ultrasound, MRI and histopathology findings. There were three such occult cases. When subsequently participating in the reader study, the two radiologists were only assigned cases on which they had not performed ground-truth marking.

The ground truth data and all the reader data were combined and exported from MedXViewer to a spreadsheet for analysis. Each lesion was assigned a unique identifier by MedXViewer, incorporating lesion-matching across different mammographic projection images. The readers' marks and interpretations captured by the software were automatically compared to the ground truth marks and diagnoses. Thus the software recorded whether a reader had successfully detected a lesion and correctly identified it as malignant or benign. To score a true positive, the reader mark had to be within the corresponding generously-sized ROI applied at ground-truth marking.

Analytical and statistical methods

The performance of the modalities was based on sensitivity and specificity and the plotting of receiver operator characteristic curves (true positives versus false positives). The analyses were conducted at the per-breast level. In order to determine population variation a Monte Carlo subsampling approach was applied to the data, where the population was sampled 20 times for a randomised subset of 30-50% of the dataset depending on the size of the data. The sensitivity and specificity of the results were then calculated for each of these Monte Carlo derived subsets. The variance in sensitivity and specificity and the confidence intervals were calculated from this population of subsets. The same values were used to plot receiver operating characteristic (ROC) curves and calculate the area under the curve (AUC) for each simulation, utilising the ROCR package

213 in R^{10}. Significance between approaches was tested using a paired two-way Student's t-test on the

214 Log normalised values.

215

Results

Sample description

Four hundred and forty six patients were recruited between March 2011 and April 2016. Figure 1 provides a recruitment flowchart and Table 1 shows the characteristics of the cases included in the retrospective multi-reader study.

Overall sensitivity

Sensitivity for breast cancer was 86.6% with FFDM (95% CI: 85.2-88.0%), 89.1% with DBT (95% CI: 88.2-90.0%), and 91.7% with FFDM+DBT (95% CI: 90.7-92.6%). Comparing the values by t-test, the differences in sensitivity for cancer between modalities were statistically significant - FFDM versus DBT: p=.004; DBT versus FFDM+DBT: p<.001; FFDM versus FFDM+DBT: p<.001.

In the reader study, there were four cases picked up by FFDM but not by either reader with DBT. The features were as follows: ill-defined mass, n=2; well-defined mass, n=1; lobulated mass with associated calcifications, n=1. There were eight cases picked up by DBT but not by either reader with FFDM. The features were as follows: spiculated mass, n=5; well-defined mass, n=1; ill-defined mass, n=1; ill-defined mass with associated calcifications, n=1.

Sensitivity according to mammographic density

For each mammogram there were two reader classifications using the BI-RADS® four-category density system[9]. The two BI-RADS® values per patient were applied to all observer-readings . The following distribution of BI-RADS® density categories was seen (n=157 breasts with cancer x 2 BI-RADS® reads; total n=314): category A (almost entirely fatty), n=23 (7%); category B (scattered areas of fibroglandular density), n=132 (42%); category C (heterogeneously dense), n=128 (41%); category D (extremely dense), n=31 (10%). Agreement between the readers on the BI-RADS® category for each patient was 62%. Variations in percentage cancer sensitivity according to BI-RADS® density category are shown in Figure 2.

For the 0-100 VAS values (observers' assessments of percentage dense area estimated for the mammogram overall), the mean of the two readers' scores was used and was applied to all breasts for the analysis. The data were divided into quartiles and the ranges for each quartile were as follows: – Q1: 4-31, Q2: 32-41, Q3: 42-63, Q4: 64-86. Variations in percentage cancer sensitivity according to VAS density are shown in Figure 3.

For volumetric percentage breast density assessed by Volpara® software, the mean of the per-image values for each patient (non-cancer-bearing breast only) was used for analysis, and the data were divided into quartiles. Ranges within the quartiles were as follows: Q1: 2.37-4.87, Q2: 4.91-7.03, Q3: 7.18-13.09, Q4: 13.15-39.05. Patients with bilateral cancer were excluded (n=5) and Volpara® data were unavailable for six patients. Variations in percentage cancer sensitivity according to Volpara® density are shown in Figure 4.

In summary, decreased sensitivity with increasing breast density was less marked with DBT than with FFDM. By all three density measures, FFDM+DBT was more sensitive than FFDM in the most dense category, whereas the advantage of DBT alone was most apparent in the third most dense category. Only the automated density assessment method (Volpara®) showed a statistically significant sensitivity increment in the most dense breasts for DBT alone: DBT 82.0% versus 74.8% for FFDM, $p<.001$.

The largest subgroup benefit detected in the study was the 10.3% sensitivity increment seen in the densest breasts according to the Volpara® measurement (85.1% with FFDM+DBT versus 74.8% with FFDM alone, $p<.001$).

Sensitivity in different tumour types

Analysing sensitivity separately for the invasive lobular (ILC) and combined non-lobular invasive cancers revealed no statistically significant differences between modalities in the lobular group: FFDM: 84.1% (95% CI: 80.4-87.8%); DBT: 85.6% (82.0-89.1%); FFDM+DBT: 87.7% (84.9-90.4%). T-test

results were: FFDM versus DBT: *p*=.55; FFDM versus FFDM+DBT: *p*=.11; DBT versus FFDM+DBT: *p*=.33.

The results for non-lobular invasive cancer sensitivity, which were overwhelmingly the larger group (136 breasts versus 19 breasts), closely reflect the overall results: (FFDM: 86.3% (95% CI: 85.6-87.1%); DBT: 89.4% (88.7-90.2%); FFDM+DBT: 90.7% (90.0-91.4%). T-test results were: FFDM versus DBT: *p*<.001; FFDM versus FFDM+DBT: *p*<.001; DBT versus FFDM+DBT: *p*=.01.

Specificity

Specificity was 81.4% with FFDM (95% CI: 80.5-82.3%), 84.6% with DBT (95% CI: 83.9-85.3%), and 79.6% with FFDM+DBT (95% CI: 79.0-80.2%). Differences were statistically significant by t-test at: FFDM vs DBT: *p*<.001; FFDM vs FFDM+DBT: *p*=.003; DBT versus FFDM+DBT: *p*<.001. Of note, in the subgroup with the highest sensitivity gain using the FFDM+DBT modality (i.e. a 10% sensitivity increment in cases with breast density in the highest Volpara® quartile) there was no specificity penalty: 87.4% with FFDM (95% CI: 85.8-89.0%), and 87.3% with FFDM+DBT (85.5-89.2%), *p*=.94.

Receiver Operating Characteristic analysis

The area under the receiver operating characteristic curve (AUC) for FFDM was 0.90; for DBT it was 0.92; for FFDM+DBT it was 0.92.

Assessment of tumour size in unifocal cancer cases

There were 214 reader measurements of unifocal malignant lesions not treated with neoadjuvant systemic therapy under the FFDM condition and 260 under the DBT condition. (The difference in numbers reflects the higher sensitivity of DBT.) Absolute agreement between reader measurements and histopathology measurements by intraclass correlation coefficient was 0.41 for FFDM (95%CI: 0.25-0.60) and 0.55 for DBT (95% CI: 0.28-0.70). The rate of overestimation of histopathological WTD by more than 5 mm was 10.3% with FFDM and 8.5% with DBT (*p*=.82). The rate of underestimation by more than 5 mm was 47.2% with FFDM and 46.2% with DBT (*p*=.50). The mean discrepancy between the readers' measurements and the histopathological WTD was identical for the two

289 modalities - a 10 mm underestimation - and Bland-Altman 95% limits of agreement were very similar
290 at +26.7 to -46.8 for FFDM and +25.1 to -45.0 for DBT. Please see Figure 5 for Bland-Altman plots.
291 Common to both modalities was a tendency for greater underestimation with increasing lesion size.
292 Invasive lobular carcinoma and lesions which included radiological microcalcifications were over-
293 represented in the top quintile of size underestimation.

Discussion

Our study contributes to the limited body of evidence evaluating the effectiveness of modern mammographic imaging in patients presenting with symptoms of possible breast cancer. While it can be argued that the importance of any single imaging modality is limited in the context of multi-modality diagnostic breast clinics, even triple assessment does not completely eliminate false negative findings in symptomatic patients[11,12] so the sensitivity performance of each individual element still matters. Furthermore, if mammography is negative in the presence of malignant clinical or sonographic findings, or if there is size discrepancy between assessment methods, there can be a tendency to resort to an expensive MRI scan, despite limited evidence supporting MRI for diagnostic problem-solving[13].

Overall sensitivity for cancer was high with all three modalities in our study (FFDM: 86.6% - FFDM+DBT: 91.7%). Given that the sensitivity of FFDM was so high, it is unsurprising that the overall gains from adding DBT were clinically relatively modest. A recent study in the symptomatic setting was conducted in China in a sample where 149 of 197 participants had BI-RADS® C or D density[7]. That study, using Hologic Selenia Dimensions equipment, did not compare FFDM with DBT alone but found that sensitivity for FFDM in their whole sample was 72% - much lower than in our sample - and for FFDM and DBT combined was 91% - similar to our value. Information on tumour size was not provided in the publication. Diagnostic studies with mixed samples of screen-detected and symptomatic lesions, using a prototype GE tomosynthesis device (GE Healthcare, Chicago, Illinois), have also shown lower FFDM sensitivity and higher sensitivity gains with DBT than were seen in our study[14,15].

Differential sensitivity by breast density in our study varied according to method of assessing density but the overall pattern is for the sensitivity gains from DBT to be more apparent in denser breasts. Again, however, our FFDM performance compares favourably with published values. In a recent study[6] comparing FFDM and DBT using Hologic Selenia Dimensions equipment in a symptomatic

population with dense breasts (BI-RADS® C or D), sensitivity for cancer was considerably lower with both modalities than for women with dense breasts in our study, at 59% for FFDM (versus a mean sensitivity in our BI-RADS® C and D cases of 79%) and 68% (versus 80%) for DBT. The mean tumour size of 23 mm in that study, compared to 32 mm in ours, may help explain the generally lower sensitivity.

In ILC, descriptively FFDM+DBT gave a 3.6% increment over FFDM alone but there were no statistically significant differences in sensitivity for ILC between modalities, possibly because there were only 19 cases of ILC in our study. A previous larger multi-reader study including screen-detected and symptomatic cases of ILC, using Hologic Selenia Dimensions equipment, found a statistically significant 15% sensitivity increment with DBT (85% versus 70%)[16]. It has also been shown that reader ratings of lesion conspicuity in ILC are higher with DBT than FFDM[17].

Specificity in our study was about 3% higher for DBT versus FFDM, with a 5% drop in overall specificity for FFDM+DBT. We think the lower specificity with the combined modalities is most likely just a function of having two tests instead of one. Tang and colleagues[7] found no difference in specificity for FFDM versus FFDM+DBT, with generally lower specificity than ours, at 72% and 71% respectively. Bian and colleagues[6] achieved higher specificity for both FFDM and DBT alone, rising to 95% for DBT, which may further explain the relatively low sensitivity values in that study. DBT has been shown to improve specificity in screening studies[18] but maximising specificity is less important in the symptomatic triple assessment clinic than in the screening of well women, especially in women with clinical suspicion of cancer as in this study.

We detected only very small differences in AUC values (0.90 for FFDM and 0.92 for DBT), similar to the study by Tang and colleagues[7], which demonstrated an improvement from 0.85 to 0.9.

Accurate estimation of tumour extent is important in guiding therapeutic decision-making. Our study detected little descriptive improvement and no statistically significant improvement in the accuracy

of measuring the size of unifocal cancers with DBT compared to FFDM. Conversely, several earlier studies have found DBT size assessment to be more accurate than FFDM[19,20], although it has also been shown that there is a greater risk of overestimation of tumour size with tomosynthesis[21,22], which was not our experience in this study. Our finding that underestimation of tumour size in ILC persists with DBT is in line with previous work[17], but that study also included only a small number of ILC cases.

Study strengths and limitations

Study strengths included the use of multiple centres and multiple readers, and the strict blinding of readers between modalities. The inclusion of multiple measures of breast density was also a strength. It could be considered a limitation that the images were read under simulated rather than real-life practice conditions, but that approach was necessary in order to conduct a robustly blinded study. Synthesised 2D images which can be used in place of standard FFDM were not available at the time of image acquisition for our study. We did not follow up the patients to ascertain false negative triple assessment cases, therefore the study assesses the relative sensitivity of the modalities. Because ours was a study of patients presenting with suspicious clinical symptoms, it does not add to the evidence base on the clinical utility of DBT in DCIS. Like others of its kind, ours was a single-vendor study and results may not be generalizable to other vendors' equipment.

Conclusions

FFDM and DBT in combination provided a small but statistically significant improvement in sensitivity for cancer in our sample of younger symptomatic patients, from 86.6% to 91.7% overall.

The greatest improvements in sensitivity, over FFDM alone, were seen with the combined modality in the densest breasts (an increment of 9% when density was measured by human-assigned area-based percentage, and of 10% when density was measured by Volpara® software).

366 The overall sensitivity improvement with combined FFDM and DBT was at the cost of a small
367 reduction in specificity, from 81.4% to 79.6%.

368 No advantage was seen for assessment of unifocal tumour size.

369 Although our study has not shown FFDM to be sufficiently inferior to mandate the replacement or
370 supplementation of FFDM with DBT for all younger women in the symptomatic clinic, where it is
371 available it does merit first-line use in the under 60 age group, particularly in women who are known
372 to have very dense breasts. If breast density is not known in advance from prior mammography, DBT
373 could be performed after negative FFDM in women with dense breasts, rather than in combination
374 at the outset.

375 The benefits of DBT should be weighed against the additional radiation dose, acquisition time,
376 reading time and data storage costs. The contribution of DBT to triple assessment in symptomatic
377 women with dense breasts needs to be reassessed in comparison with the performance of other
378 potential diagnostic tests such as increasingly-available contrast-enhanced mammography.

References

1. Pisano ED, Hendrick RE, Yaffe MJ, Baum JK, Acharyya S, Cormack JB, et al. Diagnostic Accuracy of Digital versus Film Mammography: Exploratory Analysis of Selected Population Subgroups in DMIST. Radiology [Internet]. 2008 Feb;246(2):376–83. Available from: http://pubs.rsna.org/doi/10.1148/radiol.2461070200

2. Sibbering DM, Burrell HC, Evans AJ, Yeoman LJ, Wilson ARM, Robertson JFR, et al. Mammographic sensitivity in women under 50 years presenting symptomatically with breast cancer. The Breast. 1995;4(2):127–9.

3. Häberle L, Fasching PA, Brehm B, Heusinger K, Jud SM, Loehberg CR, et al. Mammographic density is the main correlate of tumors detected on ultrasound but not on mammography. Int J Cancer. 2016;139(9):1967–74.

4. Hooley RJ, Durand MA, Philpotts LE. Advances in digital breast tomosynthesis. Am J Roentgenol. 2017;208(2):256–66.

5. Chae EY, Kim HH, Cha JH, Shin HJ, Choi WJ. Detection and characterization of breast lesions in a selective diagnostic population: diagnostic accuracy study for comparison between one-view digital breast tomosynthesis and two-view full-field digital mammography. Br J Radiol [Internet]. 2016 Jun;89(1062):20150743. Available from: http://www.birpublications.org/doi/10.1259/bjr.20150743

6. Bian T, Lin Q, Cui C, Li L, Qi C, Fei J, et al. Digital Breast Tomosynthesis: A New Diagnostic Method for Mass-Like Lesions in Dense Breasts. Breast J. 2016;22(5):535–40.

7. Tang W, Hu F-X, Zhu H, Wang Q-F, Gu Y-J, Peng W-J. Digital breast tomosynthesis plus mammography, magnetic resonance imaging plus mammography and mammography alone: A comparison of diagnostic performance in symptomatic women. Clin Hemorheol Microcirc [Internet]. 2017 Jun 10;66(2):105–16. Available from: https://content.iospress.com/articles/clinical-hemorheology-and-microcirculation/ch16242

8. Maxwell AJ, Ridley NT, Rubin G, Wallis MG, Gilbert FJ, Michell MJ. The Royal College of Radiologists Breast Group breast imaging classification. Clin Radiol [Internet]. 2009 Jun;64(6):624–7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0009926009000610

9. D'Orsi C, Sickles E, Mendelson E, Morris E, Bassett L. ACR BI-RADS Atlas. 5th ed. Reston, VA: American College of Radiology; 2013.

10. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: Visualizing classifier performance in R. Bioinformatics. 2005;21(20):7881.

11. Britton P, Duffy SW, Sinnatamby R, Wallis MG, Barter S, Gaskarth M, et al. One-stop diagnostic breast clinics: How often are breast cancers missed. Br J Cancer. 2009;100(12):1873–8.

12. Coolen A, Leunen K, Menten J, van Steenbergen W, Neven P. False-negative tests in breast cancer management. Neth J Med. 2011;69(7):324–9.

13. Lehman CD, Lee AY, Lee CI. Imaging management of palpable breast abnormalities. Am J Roentgenol. 2014;203(5):1142–53.

14. Chan HP, Helvie MA, Hadjiiski L, Jeffries DO, Klein KA, Neal CH, et al. Characterization of Breast Masses in Digital Breast Tomosynthesis and Digital Mammograms: An Observer Performance Study. Acad Radiol [Internet]. 2017;24(11):1372–9. Available from: https://doi.org/10.1016/j.acra.2017.04.016

15. Chae EY1, Kim HH1, Cha JH1, Shin HJ1 CW. Detection and characterisation of breast lesions in a selective diagnostic population: Diagnostic accuracy study for comparison between one-view digital breast tomosynthesis and two-view full-field digital mammography. Br J Radiol. 2016;13(20150743. [Epub ahead of print]):9–13.

16. Mariscotti G, Durando M, Houssami N, Zuiani C, Martincich L, Londero V, et al. Digital breast tomosynthesis as an adjunct to digital mammography for detecting and characterising invasive lobular cancers: a multi-reader study. Clin Radiol [Internet]. 2016;71(9):889–95. Available from: http://dx.doi.org/10.1016/j.crad.2016.04.004

17. Chamming's F, Kao E, Aldis A, Ferré R, Omeroglu A, Reinhold C, et al. Imaging features and conspicuity of invasive lobular carcinomas on digital breast tomosynthesis. Br J Radiol. 2017;90(1073).

18. Houssami N, Macaskill P, Marinovich ML, Hunter KE. Breast Cancer Screening Using Tomosynthesis or Mammography: A Meta-analysis of Cancer Detection and Recall. JNCI J Natl Cancer Inst [Internet]. 2018;110(9):942–9. Available from:

https://dx.doi.org/10.1093/jnci/djy121

19. Luparia A, Mariscotti G, Durando M, Ciatto S, Bosco D, Campanino PP, et al. Accuracy of tumour size assessment in the preoperative staging of breast cancer: comparison of digital mammography, tomosynthesis, ultrasound and MRI. Radiol Med [Internet]. 2013;118(7):1119–36. Available from: http://link.springer.com/10.1007/s11547-013-0941-z

20. Helal MH, Mansour SM, Zaglol M, Salaleldin LA, Nada OM, Haggag MA. Staging of breast cancer and the advanced applications of digital mammogram: what the physician needs to know? Br J Radiol [Internet]. 2017 Mar;90(1071):20160717. Available from: http://www.birpublications.org/doi/10.1259/bjr.20160717

21. Mercier J, Kwiatkowski F, Abrial C, Boussion V, Dieu-De Fraissinette V, Marraoui W, et al. The role of tomosynthesis in breast cancer staging in 75 patients. Diagn Interv Imaging [Internet]. 2015;96(1):27–35. Available from: http://dx.doi.org/10.1016/j.diii.2014.06.010

22. Marinovich ML, Bernardi D, Macaskill P, Ventriglia A, Sabatino V, Houssami N. Agreement between digital breast tomosynthesis and pathologic tumour size for staging breast cancer, and comparison with standard mammography. The Breast [Internet]. 2019 Feb 1;43:59–66. Available from: https://doi.org/10.1016/j.breast.2018.11.001

Figures

Figure 1: Participation flowchart

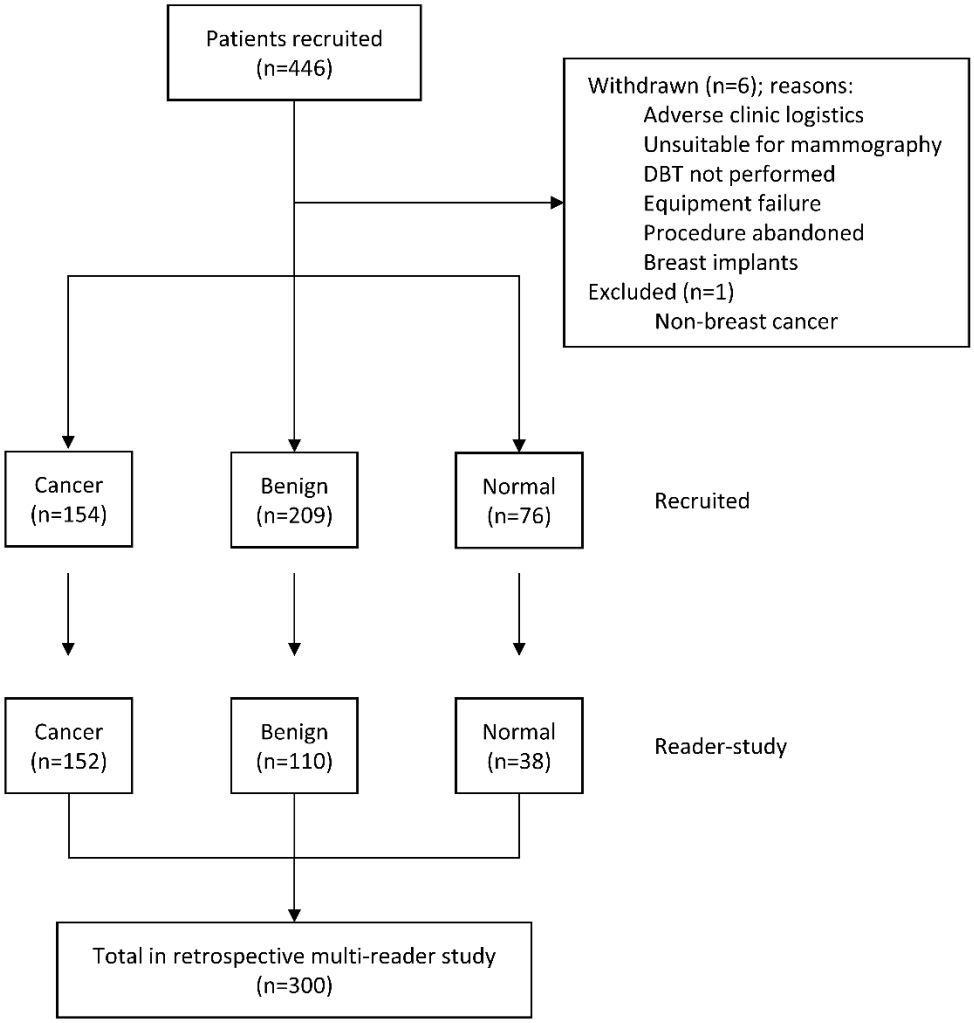


Figure 2: Sensitivity (%) according to BI-RADS® density category. (Values significantly different to FFDM at p<.05 are denoted by asterisks.)

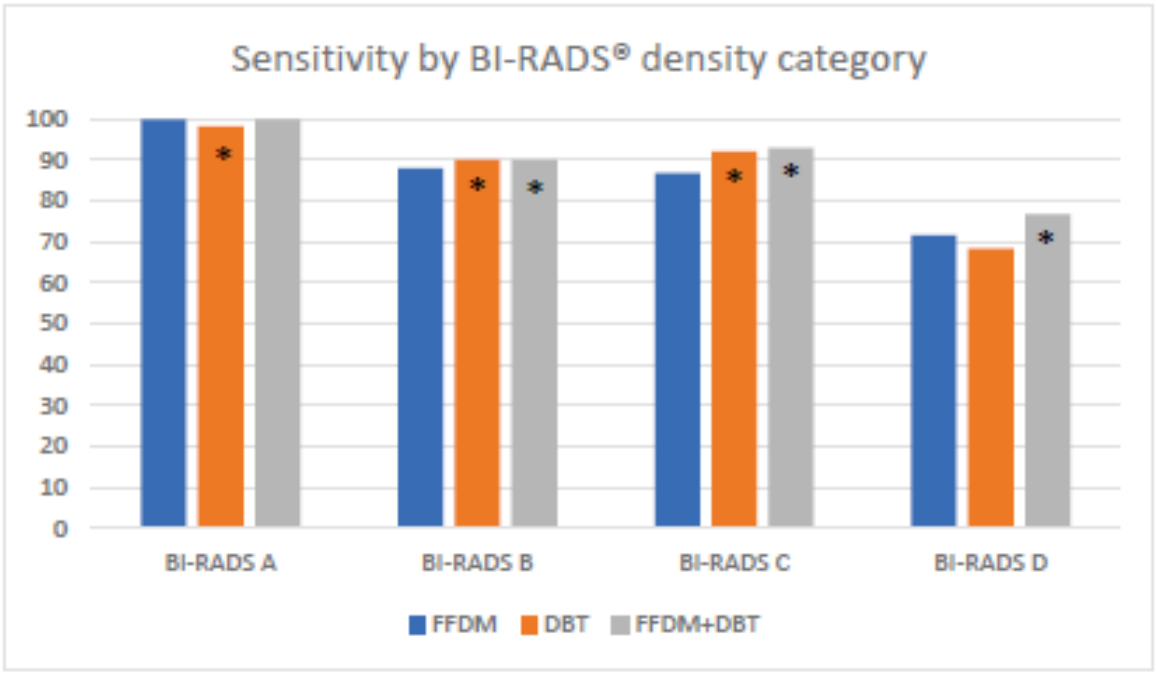


Figure 3: Sensitivity (%) according to VAS percent density quartile. (Values significantly different to FFDM at p<.05 are denoted by asterisks.)

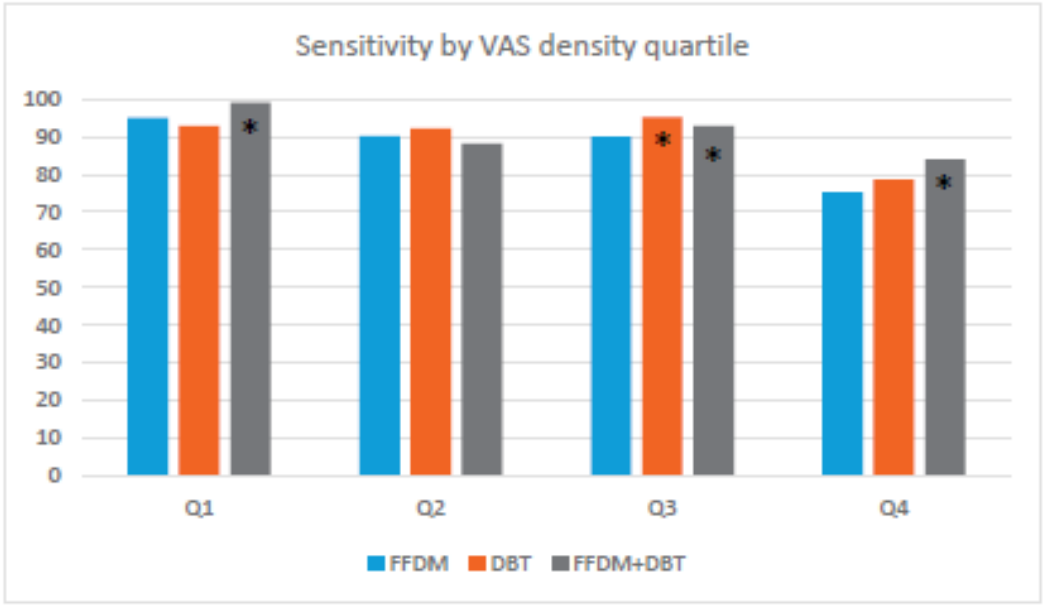


Figure 4: Sensitivity (%) according to Volpara® density quartiles. (Values significantly different to FFDM at p<.05 are denoted by asterisks.)

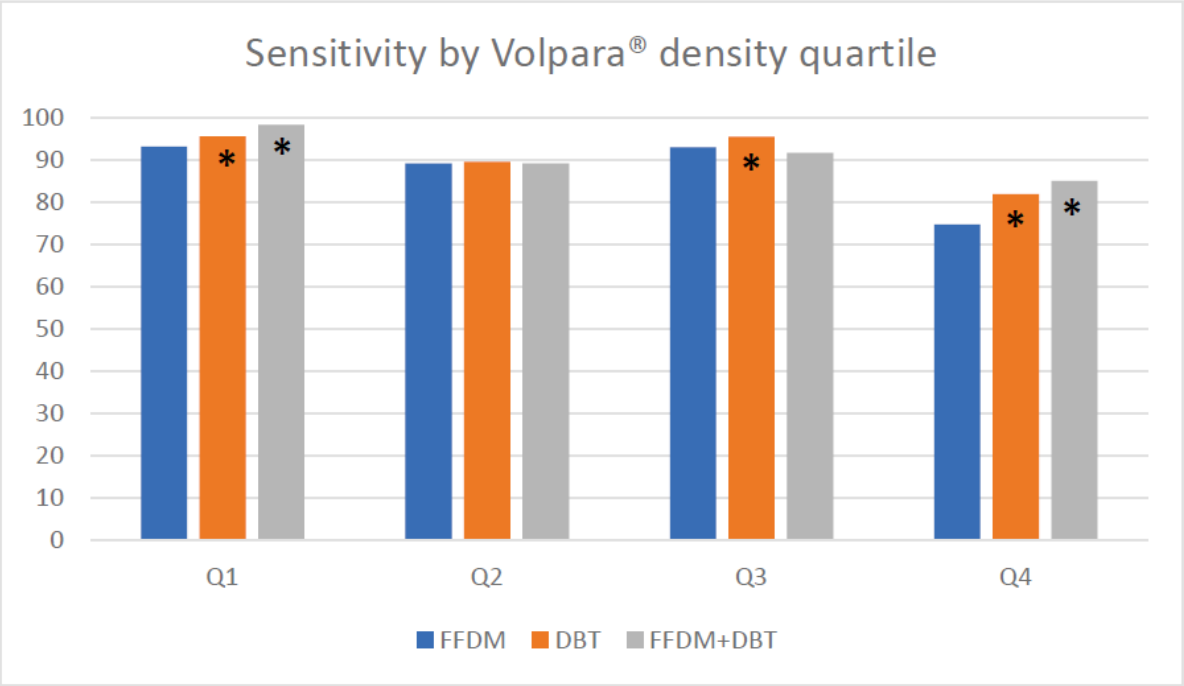
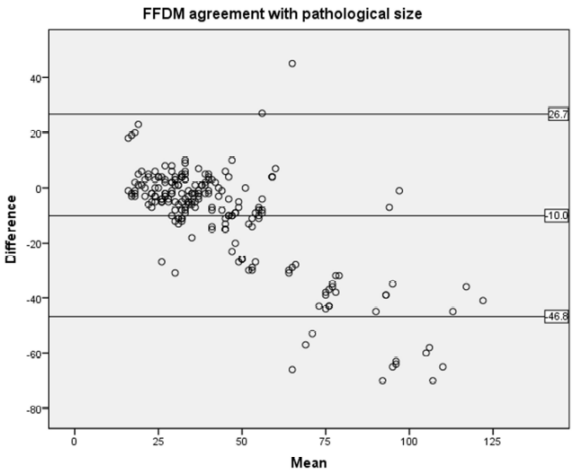
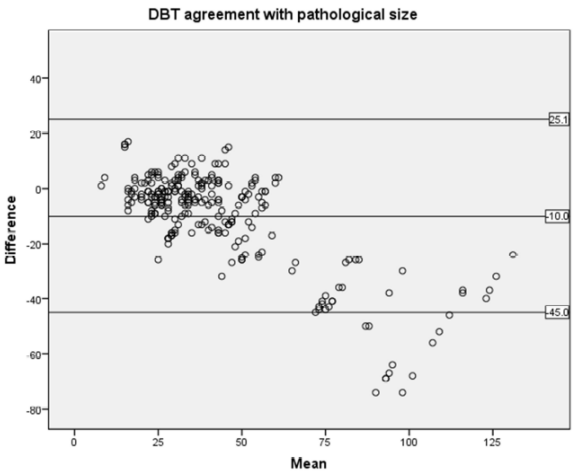


Figure 5: Bland-Altman plots for agreement between imaging tumour size and final histopathological size (unifocal only)





Tables

Table 1: Characteristics of patients included in the multi-reader study

	Patients: n=300 **Patients with cancer: n=152** **Breasts with cancer: n=157**
Mean patient age (range)	47 (24-60)
Mass as dominant radiological feature in malignancies	140/157 (89%)
Unifocal tumours	134/157 (85%)
Multifocal tumours	23/157 (15%)
Mean tumour size, unifocal breast cancers (range)	32 mm (5-95 mm)
Median tumour size, unifocal breast cancers	25 mm
DCIS	2/157 (1%)
Invasive (ductal) no special type, of which	127/157 (81%)
Grade 1	9/127
Grade 2	52/127
Grade 3	66/127
Invasive lobular carcinoma (ILC)	19/157 (12%)
Mixed ductal/lobular	2/157 (1%)
Other invasive carcinoma (Mucinous n=3, one each tubular, micropapillary, Metaplastic, malignant phyllodes	7/157 (4%)