Published in final edited form as:

Acad Radiol. 2017 November ; 24(11): 1372–1379. doi:10.1016/j.acra.2017.04.016.

Characterization of breast masses in digital breast tomosynthesis and digital mammograms: an observer performance study

Heang-Ping Chan, PhD, Mark A. Helvie, MD, Lubomir Hadjiiski, PhD, Deborah O. Jeffries, MD, Katherine A. Klein, MD, Colleen H. Neal, MD, Mitra Noroozian, MD, Chintana Paramagul, MD, and Marilyn A. Roubidoux, MD

Department of Radiology, University of Michigan, Ann Arbor

Abstract

Objectives—To compare BI-RADS assessment of lesions in two-view digital mammogram (DM) to wide-angle two-view digital breast tomosynthesis (DBT) without DM.

Materials and Methods—With IRB approval and written informed consent, two-view DBTs were acquired from 134 subjects and the corresponding DMs were collected retrospectively. The study included 125 subjects with 61 malignant (Size: 3.9–36.9 mm, median 13.4 mm) and 81 benign lesions (4.8–43.8 mm, median 12.0 mm), and 9 normal subjects. The cases in the two modalities were read independently by six experienced MQSA radiologists in a fully-crossed counter-balanced manner. The readers were blinded to the prevalence of malignant, benign, or normal cases and were asked to assess the lesions based on the BI-RADS lexicon. The ratings were analyzed by the receiver operating characteristic (ROC) methodology.

Results—Lesion conspicuity was significantly higher ($p < 0.0001$) and fewer lesion margins were considered obscured in DBT. The mean area under the ROC curve for the six readers increased significantly ($p = 0.0001$) from 0.783 (range: 0.723–0.886) for DM to 0.911 (range: 0.884–0.936) for DBT. Of the 366 ratings for malignant lesions, 343 on DBT and 278 on DM were rated as BI-RADS 4a and above. Of the 486 ratings for benign lesions, 220 on DBT and 206 on DM were rated as BI-RADS 4a and above. On average, 17.8% (65/366) more malignant lesions and 2.9% (14/486) more benign lesions would be recommended for biopsy using DBT. The inter-radiologist variability was reduced significantly.

Conclusion—With DBT alone, the BI-RADS assessment of breast lesions and inter-radiologist reliability were significantly improved compared to DM.

Keywords

Digital breast tomosynthesis; BI-RADS assessment; ROC observer study

Correspondence: Heang-Ping Chan, PhD, Department of Radiology, University of Michigan, 1500 E. Medical Center Dr., Med Inn Building C477, Ann Arbor, MI 48109-5842, Phone: 734-936-4357, chanhp@umich.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Digital breast tomosynthesis (DBT) is increasingly being used in breast imaging clinics. Three commercial DBT systems have been approved by the Food and Drug Administration (FDA) for screening and diagnostic workup and more systems are available outside the U.S. The three systems have different designs of scan parameters ranging from 15 degrees to 50 degrees. The screening protocols for the systems also differ; one system was approved for a combo mode that includes two-view DBT in combination with two-view digital mammograms (DM), the second system was approved for a craniocaudal (CC) view DM with a mediolateral oblique (MLO) view DBT, and the third system was approved for a standalone two-view DBT. A number of prospective or observational studies of clinical performance before and after implementation of DBT have been conducted to compare the combo mode with DM alone (1–10) in screening settings. All these studies found significant improvement in cancer detection rate and a reduction in the overall recall rate or a reduction in the recall rate per cancer detected. Other investigators have conducted retrospective reader studies as reviewed in (11, 12); most of these studies also revealed the potential of the combo mode, yielding higher cancer detection rate and lower recall rate compared to DM alone.

The vast majority of the studies to-date evaluated DBT as an adjunct to DM using DBT systems with a scan angle of 15°. A few studies evaluated DBT systems with larger scan angles (40° to 50°) (13–18), in which an MLO-view DBT replaced the MLO-view DM or both views of DM. The results from the studies using different modes other than the combo mode are more varied. The DM in the combo mode is also being replaced with a mammogram-like image synthesized from DBT to reduce dose and the adequacy of such approach is being investigated. DBT is still an evolving technology and its capability and limitations have not been fully explored, especially the performance of DBT of scan angles other than 15° and as a standalone modality. Continued development and studies of the impact of DBT acquisition geometry (scan angle, number of projections) and other factors on the performance of DBT are crucial to further improve its efficacy in both screening and diagnostic applications.

The purpose of our current study is to evaluate two-view DBT acquired with a prototype wide-angle (60°) DBT system as a standalone modality in the characterization of soft-tissue lesions compared to two-view DM in a retrospective observer study using receiver operating characteristic (ROC) methodology. The characteristics of lesions in terms of Breast Imaging Reporting and Data System (BI-RADS) descriptors were also compared.

2. Materials and Methods

2.1 Data set

With approval of the Institutional Review Board and written informed consent, we collected DBT of human subjects with a General Electric (GE) prototype GEN2 DBT system (GE Global Research, Niskayuna, NY). The system acquired 21 projections in 3° increments, with a total tomographic angle of 60°. To our knowledge, this is the only system that can acquire such wide-angle DBT for human subjects to-date. The subjects were recruited from

patients who had undergone diagnostic workup for a suspicious finding by screening or clinical findings. Two-view (CC and MLO) DBTs of the breast with the lesion were acquired. For each subject, the corresponding DMs were collected retrospectively from the patient archive. The time interval between the DM and the DBT ranged from 0 to 84 days (median: 13 days). Cases with microcalcifications as the only finding or cases without DM were excluded. A data set of 134 cases (age range of subjects: 29–88 years, median: 46 years) was formed, of which 125 cases contained a total of 142 lesions and 9 cases were normal.

An experienced Mammography Quality Standards Act (MQSA) qualified breast radiologist (**) marked the corresponding lesions on the DM and in the DBT volume based on all available clinical information including images and pathology reports. The radiologist also provided description of the appearance of the 142 lesions on DM, resulting in 96 masses, 14 architectural distortions, and 32 asymmetries. The data set was highly enriched with malignant cases. The pathology of 61 malignant and 74 benign lesions were proven by biopsy or fine-needle aspiration; 7 lesions were determined to be cysts by ultrasound and remained normal after 2-year follow-up. Table 1 listed the pathology of the lesions. This radiologist did not participate in the observer study.

All DBT were reconstructed with the simultaneous algebraic reconstruction technique at 0.1 mm × 0.1 mm in-plane pixel size and 1 mm slice spacing.

2.2 Observer study

The observer study was conducted with a multi-case multi-reader (MRMC) ROC methodology (19) in a fully-crossed design. Six MQSA radiologists (**, **, **, **, **, **)¹ participated as readers. At the time of the study, the six radiologists had 6 to 37 years (median 20 years) of experiences in mammography; they had not interpreted DBT in clinical settings but four had experience in reading human subject DBT (different from those in the current data set) for previous research projects.

In the observer experiment, the CC and MLO views of each modality (DBT or DM) for a breast were displayed side-by-side on two 21" 5M-pixel (2048×2560) display monitors (model EIZO SMD 21500 D), which were similar to those used in clinical FFDM review workstations. The monitors were calibrated with the DICOM grayscale standard display function. An in-house developed user interface allowed the reader to interactively apply windowing, zooming and panning to the displayed images. For DBT, the reader could scroll through the stack of DBT images of the breast volume. The interface also provided an electronic scoring sheet to record the reader's evaluation of each lesion in terms of the BI-RADS assessment of the degree of suspicion for malignancy, lesion descriptors (shape, margin, density), lesion size (the longest dimension and its perpendicular) measured with an electronic caliber, conspicuity rating (1–10, 10=most obvious), and breast density in four categories.

¹Alphabetical order, not in the same order as R1 to R6.

The cases were divided randomly into four groups of approximately equal size; each case contained two views of one breast. The two modalities of each group were then separated, resulting in a total of eight groups. To minimize effects such as reading order, memorization and learning, the reading of the modalities and cases was arranged in a counter-balanced manner for the six readers (20). Each reader read the DM groups and the DBT groups alternately and independently; three readers read a DM group first while the other three read a DBT group first. The ordering of the groups was balanced such that no group was always read in a certain order (e.g., first or last) among the six readers. For a given reader, the DM and DBT groups containing the same cases were separated as far apart as possible and with a washout time of at least one month. Before reading the study cases, each reader underwent a training session to familiarize them with the display interface, the scoring sheet, and read a set of 27 training cases of DBTs that were not part of the study cases.

The readers were informed that the data set contained both normal and abnormal cases with soft-tissue lesions. They were aware that the data set was enriched with malignant cases but were not informed of its prevalence. This study focused on BI-RADS assessment rather than detection. However, to facilitate the judgment of lesion conspicuity, the reader was asked to first search through the DBT volume and mark the lesion(s) of interest. Following the search, the lesion(s) of interest was marked on the displayed images to ensure that the same set of lesions was evaluated by all readers. The readers were blinded to all clinical information and the assessment was based exclusively on the two-view DM or DBT. The reader was asked to record their assessment of each lesion using a forced BI-RADS scale (1, 2, 3, 4a, 4b, 4c, 5) such that category 0 was not allowed. This approach is often used in retrospective observer studies so that the reader will need to provide his/her impression of the lesion based on the study images without deferring decision to additional imaging. Normal cases were included to encourage the use of the BI-RADS 1 category.

2.3 Statistical Analysis

The observer performances for assessment of the lesions on DM and DBT were analyzed with the Dorfman-Berbaum-Metz MRMCC ROC (19) software by Metz (DBM MRMCC 2.2, University of Chicago). The 7-category BI-RADS assessments formed the rating scale in the ROC analysis. The normal cases were excluded from the ROC analysis. The inter-radiologist agreement in the BI-RADS assessments was analyzed as the correlation between any two of the six radiologists and the intraclass correlation coefficient (ICC) among all six radiologists using a statistical package (SPSS Statistics version 24, IBM Corp.). The differences in the lesion size measurements and the conspicuity ratings of the corresponding lesions between DM and DBT were evaluated by the Student's two-tailed paired t-test (Microsoft Office Professional Plus 2010). A p-value of less than 0.05 was considered statistically significant.

3. Results

There are variations in the estimation of lesion characteristics and breast density by the individual radiologists. The results are therefore plotted as distributions (histograms) of the ratings or categories by all six readers as statistical representations of radiologists'

evaluation. Examples of a DBT slice in MLO view and a corresponding DM are shown in Fig. 1. The breast density categories estimated on DM are shown in Fig. 2.

The ROC curves of BI-RADS assessments for the six radiologists (R1–R6) reading DBT alone and DM alone are plotted in Fig. 3. For each radiologist, the ROC curve for DBT is higher than the ROC curve for DM. The areas under the ROC curve (AUC) are tabulated in Table 2. The ROC analysis showed that the radiologists can differentiate malignant and benign lesions more accurately in DBT than DM, with an improvement in the AUC ranging from 0.037 to 0.160, and 5 of the 6 improvements were statistically significant ($p < 0.05$). The average increase in AUC over all radiologists was 0.129 with 95% confidence interval (C.I.) of (0.069, 0.188) and a p-value of 0.0001.

The correlation of the BI-RADS assessments of the lesions between any two of the six radiologists ranged from 0.435 to 0.679 reading DM. This correlation increased to a range from 0.767 to 0.827 reading DBT. The ICC for the BI-RADS assessments among all six radiologists improved from 0.873 (95% C.I.: 0.835, 0.904) for DM to 0.948 (95% C.I.: 0.928, 0.963) for DBT, indicating that the inter-radiologists' agreement was significantly improved ($p < 0.0005$) by reading DBT instead of DM.

Category 4a is the threshold for biopsy recommendation using BI-RADS assessments. The total numbers of malignant and benign lesions that were considered to be BI-RADS 3 and below, or BI-RADS 4a and above as judged by the six radiologists in DBT and DM were summarized in Table 3. The sensitivity and specificity for classification of the breast lesions at this decision threshold were also shown. The analysis revealed that, on average, 17.8% (65/366) more of the malignant lesions on DBT would be considered suspicious than on DM. However, 2.9% (14/486) more of the benign lesions on DBT were also considered suspicious. Among the three types of lesions (mass, asymmetry, architectural distortion), DBT provided the greatest benefit in the classification of focal asymmetries, improving the sensitivity by 27.5% (28/102) and specificity by 2.2% (2/90). The loss in specificity in DBT was mainly caused by masses, for which 3.8% (15/390) of the BI-RADS assessments for benign masses changed from 3 and below to 4a and above.

The conspicuity rating of each lesion, averaged over the six radiologists, of the three types of lesions in DBT and DM are shown as scatter plots in Fig. 4(a) and 4(b) for the malignant and benign lesions, respectively. A comparison of the conspicuity ratings of the different types of lesions in DBT and DM are shown in Table 4. The lesions in DBT were significantly more conspicuous than those in DM, with a p-value < 0.0001 for the average conspicuities of all lesions and p-values < 0.0001 for individual radiologists.

Fig. 5 shows the histograms of the lesion margin descriptors in five BI-RADS categories (circumscribed, obscured, microlobulated, ill-defined, spiculated) characterized by the six radiologists. More benign lesions were seen as having circumscribed margin and more malignant lesions were seen as spiculated on DBT than on DM. The number of lesions with obscured margin on DBT was about one-third of that on DM.

The scatter plot of the lesion sizes, averaged over the six radiologists, is shown in Fig. 6. On average, the malignant lesions were measured to be 1.4 mm smaller and the benign lesions

were 1.5 mm smaller on DBT than on DM. The differences in the size measurements for both the malignant and benign lesions were statistically significant, with p-values of 0.0001 and 0.008, respectively.

4. Discussion

This study evaluated the characterization of lesions in wide-angle DBT as a standalone modality without DM. The lesions were assessed by the radiologists without diagnostic workup information, similar to screening, although the option of deferring decision to additional imaging was not allowed. However, in this retrospective study design, the prevalence of malignancy was highly enriched and a major advantage of DBT being able to reduce false positive recalls caused by overlapping dense tissue could not be taken into account. As a result, we did not estimate the cancer detection rate and recall rate as in previous studies. We compared the relative accuracy of assessing soft-tissue lesions in DBT with that in DM if radiologist finds a suspicious lesion during interpretation, which would influence the decision for detection and diagnostic workup. Our results showed a significant improvement in the AUC based on radiologists' BI-RADS assessments using DBT alone in comparison to DM, with a higher sensitivity but slightly lower specificity in differentiating malignant and benign lesions. Two previous observer studies also found that the AUC and sensitivity improved while the specificity reduced slightly (21) or stayed the same (22); the differences being that these studies used narrow-angle DBT (15°) and using the combo mode (21) or adding an MLO-view DBT to two-view DM (22), instead of reading two-view DBT alone as in our study. The consistently higher ROC curves with DBT indicates the potential that, after radiologists gain more experience in analyzing the features of lesions in DBT, they may be able to adjust their decision thresholds to improve specificity without reducing sensitivity.

On the other hand, previous studies of using DBT to replace one or two-view DM rather than the combo mode observed varied results. Some studies (15, 18, 23) found a higher cancer detection rate or better diagnostic workup but Lang et al. (18) also observed significantly higher recall rates when one-view DBT was compared to two-view DM. Other studies showed that MLO-view DBT alone (14, 17, 24) or two-view DBT alone (25) was comparable or non-inferior to two-view DM, whereas Gur et al. (26) reported that two-view DBT alone reduced the recall rate without increasing cancer detection rate. Wallis et al. (24) found that two-view DBT from a narrow-angle system (11°) was significantly better than two-view DM only for less experienced radiologists. In our study, the interpretation of two-view DBT was superior to two-view DM by all our experienced MQSA readers and five of the six improvements were statistically significant.

The ROC analysis and the ICC indicated that inter-radiologist variability in lesion assessment was significantly reduced by reading DBT compared to DM. The AUCs among the six radiologists varied only from 0.884 to 0.936 in DBT while the range was much wider from 0.723 to 0.886 in DM. The better agreement obtained with DBT was achieved despite the fact that all readers are experienced MQSA radiologists in interpretation of DM but had little or no experience in reading DBT at the time of the experiment.

The majority of the lesions were rated as having significantly higher conspicuity in DBT than in DM. This is consistent with the expectation that the overlapping tissue in DM was largely removed by DBT. The characteristics of the lesions such as the shape and margin were therefore more visible to the radiologists. Many of the lesions with obscured margin in DM were classified into other categories in DBT. More benign lesions were seen as having circumscribed margin whereas more malignant lesions were seen as spiculated. The improved conspicuity and margin characterization contributed to the improved assessment of the degrees of suspicion and the better agreement among the radiologists.

The lesion size measured on DBT was significantly smaller than that measured on DM, which may be attributed to the better lesion visibility with DBT. This may indicate that radiologists' estimate of the location of the lesion margin may tend to be more conservative to avoid missing part of the margin when the lesion is not clearly visualized.

There are limitations in this study. First, we did not collect DBT and DM during the same compression as in a combo mode. There may be variations in the image quality of the lesions in DBT and DM due to differences in compression and positioning as well as any potential changes in the breast between the two examinations. Nevertheless, it can be expected that the variations can go in either direction (better or worse) during image acquisition for either modality so that, on average over all cases, there would not be strong bias that favors one modality or the other. A second limitation is that this is a retrospective study so that the results only demonstrated the relative performance in BI-RADS assessment of soft-tissue lesions in DBT and DM without a true measurement of the sensitivity and specificity in clinical settings. A third limitation is that many of the cases were initially recalled by findings on DM during screening so that the lesion visibility might be biased in favor of DM. However, DBT still outperformed DM despite the potential disadvantage, further strengthening the results that DBT is superior to DM. A fourth limitation is that the current study focused only on the assessment of soft-tissue lesions and the interpretation of microcalcifications was not considered. Previous studies observed that detection or characterization of microcalcifications in DBT ranged from comparable or somewhat inferior to that in DM (13, 27–31), which is one of the reasons that DBT is combined with DM in a screening examination, or alternatively, a synthesized mammogram is generated from DBT to help visualize microcalcifications. Few studies have been conducted to compare the BI-RADS assessment of microcalcifications in DM to that in DBT alone or in DBT plus synthesized mammograms for DBT of different scan angles. It will be a topic of interest when the technology of generating synthetic mammogram becomes more mature and stabilized.

In summary, this is the first study comparing a very-wide-angle (60°) DBT with DM in characterization of breast lesions. It was found that radiologists can differentiate malignant and benign soft-tissue lesions significantly better in DBT alone than in DM and the agreement in BIRADS assessment among radiologists was significantly improved. It can therefore be expected that DBT can increase the chance of malignant lesions being recommended for biopsy or recalled for diagnostic workup compared to DM, and the recommendation would be more consistent among radiologists.

Acknowledgments

This work is supported by National Institutes of Health award number R01 CA151443. The content of this paper does not necessarily reflect the position of the funding agencies and no official endorsement of any equipment and product of any companies mentioned should be inferred.

Abbreviations

AUC	Area under the receiver operating characteristic curve
BI-RADS	Breast Imaging Reporting and Data System
C.I.	confidence interval
CC	craniocaudal
DBT	digital breast tomosynthesis
DM	digital mammography
ICC	intraclass correlation coefficient
MLO	mediolateral oblique
MQSA	Mammography Quality Standards Act
MRMC	multi-case multi-reader
ROC	receiver operating characteristic

References

1. 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. *European Radiology*. 2013; 23:2061–2071. [PubMed: 23553585]
2. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology*. 2013; 267:47–56. [PubMed: 23297332]
3. 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 Oncology*. 2013; 14:583–589. [PubMed: 23623721]
4. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R. Implementation of breast tomosynthesis in a routine screening practice: An observational study. *American Journal of Roentgenology*. 2013; 200:1401–1408. [PubMed: 23701081]
5. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA-Journal of the American Medical Association*. 2014; 311:2499–2507.
6. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *American Journal of Roentgenology*. 2014; 203:687–693. [PubMed: 24918774]
7. Durand MA, Haas BM, Yao XP, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology*. 2015; 274:85–92. [PubMed: 25188431]

8. Sharpe RE, Venkataraman S, Phillips J, et al. Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast tomosynthesis into a population-based screening program. *Radiology*. 2016; 278:698–706. [PubMed: 26458206]
9. Bernardi D, Macaskill P, Pellegrini M, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): A population-based prospective study. *Lancet Oncology*. 2016; 17:1105–1113.
10. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: A cohort study within the prospr consortium. *Breast Cancer Research and Treatment*. 2016; 156:109–116. [PubMed: 26931450]
11. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): A review of the evidence for use as a screening tool. *Clinical Radiology*. 2016; 71:141–150. [PubMed: 26707815]
12. Hodgson R, Heywang-Kobrunner SH, Harvey SC, et al. Systematic review of 3D mammography for breast cancer screening. *Breast*. 2016; 27:52–61. [PubMed: 27212700]
13. Andersson I, Ikeda DM, Zackrisson S, et al. Breast tomosynthesis and digital mammography: A comparison of breast cancer visibility and birads classification in a population of cancers with subtle mammographic findings. *European Radiology*. 2008; 18:2817–2825. [PubMed: 18641998]
14. Gennaro G, Toledano A, di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: A clinical performance study. *European Radiology*. 2010; 20:1545–1553. [PubMed: 20033175]
15. Svahn TM, Chakraborty DP, Ikeda D, et al. Breast tomosynthesis and digital mammography: A comparison of diagnostic accuracy. *British Journal of Radiology*. 2012; 85:E1074–E1082. [PubMed: 22674710]
16. Gennaro G, Hendrick RE, Toledano A, et al. Combination of one-view digital breast tomosynthesis with one-view digital mammography versus standard two-view digital mammography: per lesion analysis. *European Radiology*. 2013; 23:2087–2094. [PubMed: 23620367]
17. Thibault F, Dromain C, Breucq C, et al. Digital breast tomosynthesis versus mammography and breast ultrasound: A multireader performance study. *European Radiology*. 2013; 23:2441–2449. [PubMed: 23673573]
18. Lang K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: Results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. *European Radiology*. 2016; 26:184–190. [PubMed: 25929946]
19. Dorfman DD, Berbaum KS, Metz CE. Roc rating analysis: Generalization to the population of readers and cases with the jackknife method. *Investigative Radiology*. 1992; 27:723–731. [PubMed: 1399456]
20. Metz CE. Some practical issues of experimental design and data analysis in radiological roc studies. *Investigative Radiology*. 1989; 24:234–245. [PubMed: 2753640]
21. Mariscotti G, Durando M, Houssami N, et al. Digital breast tomosynthesis as an adjunct to digital mammography for detecting and characterising invasive lobular cancers: A multi-reader study. *Clinical Radiology*. 2016; 71:889–895. [PubMed: 27210245]
22. Thomassin-Naggara I, Perrot N, Dechoux S, Ribeiro C, Chopier J, de Bazelaire C. Added value of one-view breast tomosynthesis combined with digital mammography according to reader experience. *European Journal of Radiology*. 2015; 84:235–241. [PubMed: 25467641]
23. Waldherr C, Cerny P, Altermatt HJ, et al. Value of one-view breast tomosynthesis versus two-view mammography in diagnostic workup of women with clinical signs and symptoms and in women recalled from screening. *American Journal of Roentgenology*. 2013; 200:226–231. [PubMed: 23255766]
24. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: High-resolution x-ray imaging observer study. *Radiology*. 2012; 262:788–796. [PubMed: 22274840]
25. Teertstra HJ, Loo CE, van den Bosch MAAJ, et al. Breast tomosynthesis in clinical practice: Initial results. *European Radiology*. 2010; 20:16–24. [PubMed: 19657655]

26. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: Observer performance study. *American Journal of Roentgenology*. 2009; 193:586–591. [PubMed: 19620460]
27. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: Initial experience in 98 women with abnormal digital screening mammography. *American Journal of Roentgenology*. 2007; 189:616–623. [PubMed: 17715109]
28. Spangler ML, Zuley ML, Sumkin JH, et al. Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: A comparison. *American Journal of Roentgenology*. 2011; 196:320–324. [PubMed: 21257882]
29. Kopans D, Gavenonis S, Halpern E, Moore R. Calcifications in the breast and digital breast tomosynthesis. *Breast Journal*. 2011; 17:638–644. [PubMed: 21906207]
30. Tagliafico A, Mariscotti G, Durando M, et al. Characterisation of microcalcification clusters on 2D digital mammography (FFDM) and digital breast tomosynthesis (DBT): Does DBT underestimate microcalcification clusters? Results of a multicentre study. *European Radiology*. 2015; 25:9–14. [PubMed: 25163902]
31. Clauser P, Nagl G, Helbich TH, et al. Diagnostic performance of digital breast tomosynthesis with a widescan angle compared to full-field digital mammography for the detection and characterization of microcalcifications. *European Journal of Radiology*. 2016; 86:2161–2168.

Figure 1.

Example of DBT slices and corresponding DM in MLO view: (a) DBT slice at a depth of 14 mm from the compression paddle showing an invasive ductal carcinoma, (b) DBT slice at 40 mm from the compression paddle showing a second invasive ductal carcinoma, (c) corresponding DM. The lesions are marked by a white arrow on each image.

Figure 2.
The breast density categories from the six radiologists reading the digital mammograms (DM) of the breasts in the data set.

Figure 3.
Comparison of the ROC curves fitted to the RI-RADS assessment ratings for the six radiologists reading the two modalities independently.

Figure 4.

Scatter plot of conspicuity ratings, averaged over the six radiologists, of the (a) malignant and (b) benign lesions seen in DBT alone and DM alone. The conspicuity of the lesions was rated on a 10-point scale (10=most conspicuous). All but a small fraction of masses were found to be more visible in DBT than in DM. The improvement in conspicuity was statistically significant for each type of lesions.

Figure 5. Lesion margin characteristics assessed by six radiologists using BI-RADS descriptors by reading on (a) DBT, (b) DM.

Figure 6.

Scatter plot of lesion sizes, averaged over the six radiologists, of the lesions seen in DBT and DM. Most of the lesions were measured to be smaller in DBT. The lesion size was measured as the longest diameter seen on the images.

Table 1

The pathology of the malignant or benign lesions included in the study.

Lesion type	Pathology	Number
Benign	Fibroadenoma	43
	Fibrocystic change	8
	Fat necrosis	2
	Lymph node	2
	Usual intraductal hyperplasia	4
	Atypical ductal hyperplasia and atypical lobular hyperplasia	1
	Pseudoangiomatous stromal hyperplasia	1
	Cyst	11
	Hematoma	2
	Benign tissue	7
Malignant	Invasive ductal carcinoma	31 *
	Invasive lobular carcinoma	12 *
	Invasive carcinoma with ductal and lobular features	14 *
	Adenocarcinoma	1
	Invasive Tubular carcinoma	1
	Mucosa-associated lymphoid tissue (MALT) lymphoma	2

* 28 of the invasive ductal or lobular carcinomas also had ductal carcinoma in situ and/or lobular carcinoma in situ.

Table 2

The areas under the ROC curves for the six radiologists reading the two modalities (DBT, DM) and rating the degree of suspicion by BI-RADS assessment categories (1, 2, 3, 4a, 4b, 4c, 5).

Reader	Area under the ROC curve			p-value
	DBT	DM	Difference (DBT-DM)	
R1	0.916	0.769	0.147	0.0018
R2	0.902	0.762	0.140	0.0003
R3	0.906	0.770	0.137	0.0011
R4	0.923	0.886	0.037	0.2567
R5	0.884	0.723	0.160	0.0003
R6	0.936	0.787	0.149	<0.0001
Mean	0.911	0.783	0.129	
Standard deviation	0.020	0.034		
95% Confidence Interval			(0.069, 0.188)	
p-value			0.0001	

Table 3

The BI-RADS assessments by the six radiologists in the two modalities (DBT, DM) for malignant or benign lesions grouped into three types (architectural distortion, asymmetry, and mass) based on their appearance on DM. The number of lesions in each type can be found in Table 4. The total number of ratings was 366 (61×6) for malignant lesions and 486 (81×6) for benign lesions. The sensitivity and specificity for the classification of malignancy are determined with a decision threshold of BI-RADS 4a and above. For malignant lesions, positive change to BI-RADS 4a and above is desirable. For benign lesions, positive change to BI-RADS 4a and above is undesirable.

Ratings	DBT			DM			Change (DBT-DM)
	Malignant	Benign		Malignant	Benign		
BI-RADS 3 and below	Architectural distortion	1	0	22	1	-21	-1
	Asymmetry	8	59	36	57	-28	+2
	Mass	14	207	30	222	-16	-15
	Total	23	266	88	280	-65	-14
BI-RADS 4a and above	Architectural distortion	77	6	56	5	+21	+1
	Asymmetry	94	31	66	33	+28	-2
	Mass	172	183	156	168	+16	+15
	Total	343	220	278	206	+65	+14
Classification	Sensitivity	Specificity		Sensitivity	Specificity	Percentage change	Percentage change
	93.7% (343/366)	54.7% (266/486)		76.0% (278/366)	57.6% (280/486)	+17.8% (65/366)	-2.9% (14/486)

Table 4

Mean conspicuity ratings of the three types of lesions: architectural distortion, asymmetry, and mass, averaged over the 6 radiologists in the two modalities. The conspicuities in DBT are significantly higher than those in DM for all types of lesions with p-values from two-tailed paired-t test much smaller than 0.05. Rating scale: 1–10, 10=most conspicuous.

	Malignant			Mean conspicuity			Benign			Mean conspicuity		
	N	DBT	DM	p-value	N	DBT	DM	p-value	N	DBT	DM	p-value
Architectural distortion	13	5.8	3.3	0.0002	1	4.0	2.5	-				
Asymmetry	17	5.2	3.1	0.00001	15	5.4	3.3	0.00001				
Mass	31	6.7	5.9	0.0011	65	7.4	5.9	<<0.0001				
Total	61	6.1	4.6	<<0.0001	81	7.0	5.4	<<0.0001				