

Published in final edited form as:

Acad Radiol. 2012 February ; 19(2): 166–171. doi:10.1016/j.acra.2011.10.003.

Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study

David Gur, ScD¹, Margarita L. Zuley, MD², Maria I. Anello, DO², Grace Y. Rathfon, MD², Denise M. Chough, M.D.², Marie A. Ganott, M.D.², Christiane M. Hakim, M.D.², Luisa Wallace, MD², Amy Lu, MD², and Andriy I. Bandos, PhD³

¹University of Pittsburgh, Department of Radiology, Radiology Imaging Research, 3362 Fifth Avenue, Pittsburgh, PA 15213

²Department of Radiology, Magee-Womens Hospital, 300 Halket Street, Pittsburgh, PA 15213

³University of Pittsburgh, Graduate School of Public Health, Department of Biostatistics, 130 DeSoto Street, Pittsburgh, PA 15261

Abstract

Rationale and Objectives—Retrospectively compare interpretive performance of synthetically reconstructed two-dimensional images in combination with DBT versus FFDM plus DBT.

Materials and Methods—Ten radiologists trained in reading tomosynthesis examinations interpreted retrospectively, under two modes, 114 mammograms. One mode included the directly acquired FFDM combined with DBT and the other, synthetically reconstructed projection images combined with DBT. The reconstructed images do not require additional radiation exposure. We compared the two modes with respect to “sensitivity”, namely recommendation to recall a breast with either a pathology proven cancer (n=48) or a high risk lesion (n=6); and “specificity”, namely no recommendation to recall a breast not depicting an abnormality (n=144) or depicting only benign abnormalities (n=30).

Results—The average sensitivity for FFDM with DBT was 0.826 versus 0.772 for synthetic FFDM with DBT (difference=0.054, p=0.017 and p=0.053 for fixed and random reader effect, respectively). The fraction of breasts with no, or benign, abnormalities recommended to be recalled were virtually the same: 0.298 and 0.297 for the two modalities, respectively (95% confidence intervals for the difference CI= -0.028, 0.036 and CI = -0.070, 0.066 for fixed and random reader effects, correspondingly). Sixteen additional clusters of micro-calcifications (“positive” breasts) were missed by all readers combined when interpreting the mode with synthesized images versus FFDM.

Conclusion—Lower sensitivity with comparable specificity was observed with the tested version of synthetically generated images versus FFDM, both combined with DBT. Improved synthesized images with experimentally verified acceptable diagnostic quality will be needed to eliminate double exposure during DBT based screening.

© 2011 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

Correspondence: David Gur, ScD, University of Pittsburgh, Department of Radiology, Radiology Imaging Research, 3362 Fifth Avenue, Pittsburgh, PA 15213, Phone: 412-641-2513, Fax: 412-641-2582, gurd@upmc.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.

Keywords

digital breast tomosynthesis; mammography; observer performance; recall; synthetic 2D breast imaging

Introduction

Digital breast tomosynthesis (DBT) has been investigated for several years for the possible use of this technology, among others, in screening for the early detection of breast cancer [1–7]. The largest benefit demonstrated to date in this context is the possibility of significantly reducing recall rates, with some indications that observer performance in detecting specific “mass like” abnormalities could also be improved, albeit to a lesser extent [7, 8]. Recently, the Food and Drug Administration (FDA) approved the use of tomosynthesis in breast cancer screening [9]. However, DBT in combination with full field digital mammography (FFDM), which is the considered practice as presented to and approved by the FDA, requires approximately doubling the radiation dose to the breast being imaged. The primary reason for the practice of this combined procedure is the concern that some abnormalities, in particular micro-calcification clusters, will not be as readily and as easily detected and/or correctly interpreted on the tomosynthesis image sets as on conventional FFDM projection images [10]. With the knowledge that there are reasonably simple ways to reconstruct 2D projection images, as well as 3D image series, from the information acquired during a DBT data acquisition procedure, double exposure could potentially be eliminated during the combined procedure if it can be demonstrated that the 2D images reconstructed from DBT datasets results in satisfactory image quality. As a result, radiation dose would be reduced by approximately 50% to a comparable level commonly used in 2D alone mammographic procedures. Before we unilaterally and widely implement a “double dose” DBT in the screening environment, we need to assess the possible use of synthetically reconstructed 2D images during the interpretation. Therefore, we performed a preliminary retrospective observer performance study as described herein for this very purpose.

Materials and Methods

A group of FFDM and DBT examinations performed between 2008 through 2009 on 118 women ranging in age from 36 to 77 years (mean age 51 years \pm 8.7 years) were specifically selected for this study. Selection was based on the availability of a matched 2D (FFDM)/3D (DBT) image set and a predefined set of findings as a result of the final interpretation and follow up status verification. Examinations were excluded when the findings of interest were judged to be quite obvious to detect and interpret regardless of the viewing mode. All women were recruited under institutional review board approved protocols with written informed consent when they arrived at our breast imaging facility for a either screening, a diagnostic workup, or a biopsy procedure. Images were acquired with a combination protocol in which conventional FFDM is acquired first followed by a tomosynthesis acquisition technique. During the combined FFDM/DBT acquisition the breast is compressed in a conventional manner, and a FFDM image is obtained and then the x-ray tube moves along a limited arc allowing for 15 low dose images (“frames”) to be acquired rather than the single image acquired during the FFDM acquisition. After acquisition, the data from the frames are used to reconstruct 1 millimeter thick slices, the number of which varies depending on the thickness of the compressed breast. The radiation dose associated with the series of low-dose projection images is approximately the same as that of a projection mammogram with average mid- breast dose of approximately 2mGy per view.

In addition to the 3D reconstructed image set (DBT), a “synthetic” 2D image can be generated from each set of tomosynthesis slices, which is basically meant to simulate a conventional 2D FFDM image. The synthetic 2D image is created by summing and filtering the stack of reconstructed tomosynthesis slices. The image processing used is designed to generate synthesized 2D images that “look and feel” as a conventional FFDM while enhancing the visibility of calcifications and glandular tissue; hence, enabling the radiologist to use the synthetic 2D image during the interpretation as he/she would a conventional FFDM image, namely for comparison to priors, identification of mass like abnormalities and/or distortions, assessment of left/right breast asymmetry, and the detection of micro-calcification clusters. This general image processing approach was developed by Hologic Inc. (Bedford, CT), and a more detailed description of the method is described elsewhere [11]. The primary interest in these images as related to this work lies in the fact that reconstructing the synthetic 2D images from the 3D datasets does not require any additional radiation exposure. This acquisition protocol and processing procedure resulted in registered four view mammograms for all participants and each view included an actually acquired projection view (2D), a DBT (3D) reconstructed image set, and a “synthetically” reconstructed projection (2D) image [figure 1]. Two radiologists who were aware of the actual findings, had other source documents, and who did not participate as readers reviewed all the cases prior to commencement of the reader study to verify that all findings of interest were depicted (“visible”) on all image sets.

Ten board-certified, MQSA (Mammography Quality Standards Act) qualified radiologists with breast imaging experience ranging from 3 to 32 years volunteered to participate as readers in the study. The radiologists were trained in the interpretation of tomosynthesis examinations over the last five years through participation in previous reader studies and a review of different case sets with diagnostic outcome. Specific to this study, the radiologists reviewed and rated a group of positive and negative cases under the study conditions followed by a provision of the “diagnostic truth”. The radiologists retrospectively interpreted an enriched set of 114 mammograms in a fully crossed, mode balanced study. Namely, five readers first read the originally acquired FFDM images with DBT and the other five readers first read the synthetic 2D images with DBT, and after finishing their first assigned mode, and after a pre-determined time period of four weeks, the readers switched to read their remaining mode. The readers had no knowledge of the specific study objectives. Prior to commencement of the readings the radiologists received a detailed “Instructions to Observers” document defining the task at hand and the protocol was described and tested on four “test” cases in individualized introductory sessions. The “Instructions to Observers” document defined the type of examinations used in this study and provided the general set up and protocol for reviewing and rating the examinations. The document also informed readers that computer aided detection (CAD) would not be provided and no prior FFDM examinations would be provided for comparison; hence, the readers were to assume the screening examination is the woman’s baseline (first) exam. Radiologists reported/scored their breast based interpretation and recommendations using a screening Breast Imaging-Reporting and Data System (BI-RADS) rating scale (0, 1, or 2) when viewing four view combination mammography studies under the two reading modes. One mode included the original, directly acquired FFDM images combined with DBT and the other mode included the synthetically reconstructed 2D projection images combined with DBT. Readers were provided with a scoring form that included screening BI-RADS to be marked for the right and left breasts separately. Readers were to circle a screening BI-RADS rating (0, 1, or 2) for each breast. A case number that was to be matched with the case number viewed on the workstation was provided on each scoring form. No location information regarding abnormalities in question within each breast was ascertained. There was a four week interval between reading modes and most readers were able to complete a mode in one session.

All reviews and ratings were performed on a Hologic Inc. (Bedford, CT) modified SecurView mammography workstation. This is a research workstation which is PC based and includes two 5 megapixel LCD displays with a mammography workflow keypad. The system includes tools for magnification, zoom, contrast adjustment as well as a drag and drop image display. The image display software allows for the viewing of 1, 2 or 4 images per display for each monitor, so up to 8 images could be displayed simultaneously and the system allows for reader selected display and manipulation of the tomosynthesis data sets.

Description of the dataset

Each examination consisted of four views (craniocaudal (CC) and mediolateral oblique (MLO) views of each breast). From the 118 cases selected for the study, four cases (one negative, one benign, and two verified cancers) were selected for presentation and protocol testing during the individualized introductory session before the commencement of the actual readings and these examinations were excluded from the analyses. Of the 114 remaining cases, or 228 breasts, used in the study, 40 (35.1%) examinations were verified as bilaterally “negative”, 26 (22.8%) examinations depicted benign findings only, 46 (40.3%) examinations had verified cancer (pathology), and two (1.8%) examinations were verified as high risk. Of the 46 cancer cases, 2 (4%) cases had cancer in both breasts and 4 (9%) cases had verified high risk in the contra-lateral breast to that with a verified cancer. Negative examinations were verified twice (once during case selection and again when readings were completed) using follow-up (subsequent) negative imaging examinations prior to commencement of the analysis. As all analyses were performed by breast, Table 1 provides the distribution of negative, benign, and positive breasts by number and type of findings included in the study. We acknowledge that classifying papillomas without atypia as high risk lesions remains controversial; however, both cases verified with papillomas in this set also had breast cancer in the contra-lateral breast and therefore were eventually surgically excised. For the purpose of this study we assumed that these papillomas should be identified. Of the 48 verified cancers, 30 (62.5%) depicted a mass, 12 (25.0%) were depicted as microcalcifications, and six (12.5%) depicted both a mass and microcalcifications. Three of the high risk lesions depicted microcalcifications alone and the other three depicted a mass. The average size (maximum dimension -pathology based) of the cancers was 2.2 cm \pm 1.4 cm and ranged from 0.11 cm to 5.40 cm. The subjective breast density ratings distribution (BIRADS) of these cases as provided during the original clinical interpretation of the FFDM, were 3/114 (2.6%), 32/114 (28.1%), 70/114 (61.4%), and 9/114 (7.9%) for tissue density almost entirely fat, scattered fibro-glandular density, heterogeneously dense, and extremely dense, respectively. The subjective breast density ratings for the cancer cases were 2/46 (4.3%), 18/46 (39.1%), 25/46 (54.3%), and 1/46 (2.2%) for tissue density almost entirely fat, scattered fibro-glandular density, heterogeneously dense, and extremely dense, respectively.

Data Analysis

We compared the two modes with respect to breast based “sensitivity”, namely a breast with either a pathologically proven cancer or a high risk lesion being recalled; and “specificity”, namely a breast with no abnormality or only benign abnormalities not recalled. For the fixed-reader inferences we used the generalized linear mixed model (proc genmod, SAS v. 9.2) where we accounted for the correlation between the same examinations of the same cases (read under different modalities and/or by different readers) and between the assessments of different breasts of the same patient. For the random-reader inferences we used bootstrap-percentile confidence intervals based on 10,000 bootstrap samples constructed by re-sampling independently “positive” patients with cancer/high risk abnormalities and women with benign/no abnormalities (“negative”) and readers.

Results

Table 2 summarizes the estimates of recall rates for the “positive” breasts with cancer/high risk abnormalities and “negative” breasts with no or benign abnormalities. The reader-averaged sensitivity for actual FFDM with DBT was 82.6% versus 77.2% for synthetic FFDM with DBT. The increase in sensitivity when using original FFDM with DBT was observed for 9 out of 10 radiologists. Reader-averaged false positive recall rates were 29.8% and 29.7% for the two modalities with 5 out of 10 radiologists having higher specificity with actual FFDM plus DBT.

The use of synthetic rather than original FFDM images resulted in a decrease in sensitivity of 0.054 with $p=0.017$ for fixed reader effects and $p=0.053$ for random reader effects, while maintaining approximately the same false positive rate with the 95% confidence intervals for the difference of $(-0.028, 0.036)$ and $(-0.070, 0.066)$ under the fixed and random reader effects, correspondingly. A post study repeat analysis after excluding the 6 breasts with only high risk findings showed the same trend, namely difference in sensitivity > 0.04 with the same specificity for both modes, albeit the statistical significance is lost ($p>0.05$).

Despite the fact that all abnormalities of interest were depicted (“visible”) on the respected image sets, under the synthetic imaging based interpretation mode there were a total of 16 additional micro-calcification related abnormalities (“positive” breasts), or on average 1.6 per reader (over a combined total of 5 cases), that had been missed (or detected but interpreted incorrectly) when interpreting the mode with synthesized images as compared with actual FFDM.

Discussion

If DBT is to be widely used in screening mammography many issues related to ergonomics, workflow, efficiency, comparison to prior studies and CAD, among others, will have to be investigated, well understood, and carefully addressed (12). Unfortunately, investigations of these issues are limited when examined in retrospective studies. However, the issue of radiation dose can be initially investigated primarily through retrospective observer performance studies. In the study presented here we attempted to assess whether or not one version of synthetically reconstructed projection (2D) images result in sufficient image quality to enable the substitution of actually acquired FFDM images and thereby potentially reducing radiation dose of the procedure by approximately one half. We found that, while the reconstructed images we used in this study were of reasonably high quality, the images were not adequate as a substitute to the original FFDM images when used in combination with DBT in that sensitivity decreased significantly ($p<0.05$). However, we note that 2D reconstruction approaches from 3D datasets continue to evolve with new approaches/schemes that may further improve the diagnostic quality of synthetic images. At this point, we believe that, unless proven otherwise, DBT alone should not be viewed as acceptable for replacing projection images for the purpose of detecting micro-calcification clusters as this approach may result in a loss in sensitivity.

Our study has several limitations. First, this is a preliminary study using an advanced version of the reconstruction scheme at the time, but improved versions may be developed in the future. However, since, prior to the commencement of any interpretations, the processing version to be used in the reader study needed to remain constant, negative results should not be taken as conclusive. Obviously, newer (“perceptually improved”) versions of synthetically reconstructed 2D images will have to be tested independently. Second, neither prior examinations nor CAD results were available for viewing and interpretations. This may have affected results, in particular in terms of the differences in the detection of micro-

calcification clusters. Third, despite the similarity in presentation and the fact that we did not reveal the underlying objective of the study, several of the radiologists noted that in one mode the images appeared somewhat different than the displayed version they are used to in terms of “details and sharpness”. The radiologists made comments that the reconstructed images (without specifically knowing what these were) were “diagnostic” in quality, but were of different “look and feel” as well as of somewhat lower quality. Clearly, the actual results were in concordance with their subjective observations in this regard. Fourth, this was a retrospective laboratory study and actual readings in the clinic may yield different results. Last, in an attempt to mimic clinical ratings, the study used screening BIRADS for scoring cases by breast which we effectively analyzed as a binary recall/no recall response. As a result, we compared “operating points”, namely recommendation to recall predefined “actually positive” versus “actually negative” cases, but we cannot estimate the entire performance curve. In the actual clinical environment readers could be more or less aggressive in using specific response categories, affectively shifting their performance along a performance curve. However, even in this case, the fact that we observed a decrease in sensitivity accompanied by the absence of a meaningful change in specificity indicates that differences in performance levels under these modalities are unlikely to be caused by a shift along the same performance (e.g. ROC) curve.

Conclusion

Moderately higher sensitivity with virtually the same specificity was observed when interpreting actual FFDM images as compared with a current version of synthetically generated 2D projection images, both combined with DBT. Overall, performance differences between using originally acquired and synthetically reconstructed images were statistically significant ($p < 0.05$) for fixed readers effect. Improved synthesized 2D images may be available in the near future. However, whether or not the improved images are of acceptable diagnostic quality, thereby eventually leading to the possible elimination of double exposure during DBT based screening, will have to be tested independently.

Acknowledgments

This work is partially supported by Grants CA143019 and CA144055 from the National Cancer Institute, National Institutes of Health to the University of Pittsburgh.

References

1. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. *Radiology*. 1997; 205:399–406. [PubMed: 9356620]
2. Moore, RH.; Kopans, DB.; Rafferty, EA., et al. Initial callback rates for conventional and digital breast tomosynthesis mammography comparison in the screening environment. Presented at RSNA. 2007. http://rsna2007.rsna.org/rsna2007/v2007/conference/event_display.cfm?id=66601&em_id=5004025
3. Hologic submission to FDA panel for Approval of tomosynthesis. September 24, 2010 <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/RadiologicalDevicesPanel/UCM226757.pdf>
4. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *Am J Roentgenol*. 2007; 189:616–623. [PubMed: 17715109]
5. Smith A, Niklason L, Jing Z. Performance of breast tomosynthesis as an adjunct imaging modality to digital mammography. (Abs). *European Radiology*. February.2008 18(1):SS 202 B-042, 150. ECR 2008 book of abstracts.

6. Smith A, Rafferty E, Niklason L. Breast tomosynthesis reduces radiologist performance variability compared to digital mammography. (Abs). European Radiology March 2009, ECR 2009 book of abstracts. 19(1):S151. SS202 B-043.
7. Gur D, Abrams GS, Chough DM, et al. Digital Breast Tomosynthesis– An Observer Performance Study. *AJR Am J Roentgenol*. 2009 Aug; 193(2):586–91. [PubMed: 19620460]
8. Gur D, Bandos AI, Rockette HE, Zuley ML, Sumkin JH, Chough DM, Hakim CM. Localized Detection and Classification of Abnormalities on FFDM and Tomosynthesis Examinations Rated under an FROC Paradigm. *AJR Am J Roentgenol*. 2011 Mar; 196(3):737–41. [PubMed: 21343521]
9. U.S. Department of Health & Human Services. U.S. Food and Drug Administration. Medical Devices, Products and Medical Procedures, Device Approvals and Clearances, Recently Approved Devices.
<http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm246400.htm>
10. Spangler ML, Zuley ML, Sumkin JH, Ganott MA, Chough DM, Shah R, Gur D. Detection and classification of calcification on digital breast tomosynthesis and 2D digital mammography: A comparison. *AJR Am J Roentgenol*. 2011 Feb; 196(2):320–4. [PubMed: 21257882]
11. Ruth, C.; Smith, A.; Stein, J. US Patent. 7760924. System and Method for Generating a 2D Image from a Tomosynthesis Data Set.
<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnethtml%2FFPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=7760924.PN.&OS=PN/7760924&RS=PN/7760924>
12. Baker, JA.; Gur, D.; Rafferty, EA. Hot Topics: Digital Tomosynthesis: Is this an important new breast imaging technique?. Presented at RSNA 2010. SH40;

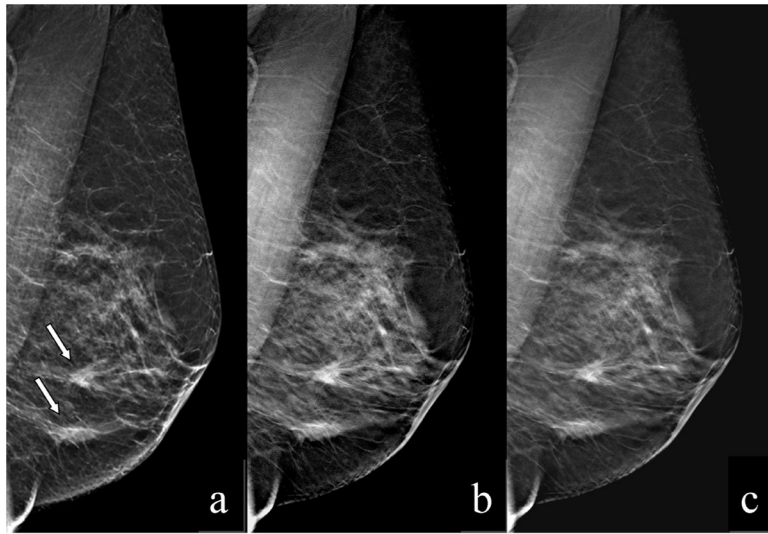


Figure 1. MLO images of the left breast of a 59YO woman depicting 2 masses, both were pathology verified as IDC and DCIS. The actually ascertained FFDM (a), the synthetically reconstructed projection image (2D) from the 3D dataset (b), and one slice (1mm thick) from the tomosynthesis image set (c), are shown.

Table 1

Distribution of breasts with verified positive, benign, and negative findings

Outcome	Finding	Number of Breasts
Cancer		
	IDC only	8
	IDC and DCIS	15
	IDC and ILC	1
	IDC, DCIS, and HR	4
	IDC and metaplastic carcinoma	1
	IDC and HR	3
	DCIS only	6
	DCIS and HR	7
	ILC only	3
High Risk		
	ADH	1
	ADH and LCIS	1
	Papilloma	2
	ALH	2
Verified Benign		30
Negative		144
Total		228

IDC - invasive ductal carcinoma

DCIS - ductal carcinoma in situ

ILC - infiltrating lobular carcinoma

HR - high risk

ADH - atypical ductal hyperplasia

ALH - atypical lobular hyperplasia

LCIS - lobular carcinoma in situ

Table 2
Breast based recommendations (BIRADS) for recalling “positive” and “negative” breasts for additional workup.

Reader	Actually acquired FFDM + DBT				Synthetically reconstructed FFDM +DBT			
	Breasts with none or only benign abnormalities		Breasts with cancer or high risk abnormalities		Breasts with none or only benign abnormalities		Breasts with cancer or high risk abnormalities	
	Recall rate	N	Recall rate	N	Recall rate	N	Recall rate	N
<i>1</i>	35.1	174	85.2	54	23.6	174	81.5	54
<i>2</i>	21.8	174	75.9	54	22.4	174	70.4	54
<i>3</i>	19.0	174	81.5	54	26.4	174	77.8	54
<i>4</i>	36.2	174	92.6	54	37.9	174	81.5	54
<i>5</i>	52.3	174	92.6	54	40.8	174	85.2	54
<i>6</i>	24.1	174	79.6	54	21.3	174	70.4	54
<i>7</i>	23.6	174	81.5	54	46.0	174	77.8	54
<i>8</i>	36.2	174	74.1	54	41.4	174	79.6	54
<i>9</i>	27.6	174	81.5	54	19.0	174	74.1	54
<i>10</i>	22.4	174	81.5	54	17.8	174	74.1	54
<i>Average</i>	29.8		82.6		29.7		77.2	