HHS Public Access

Author manuscript

Acad Radiol. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

Acad Radiol. 2015 December ; 22(12): 1477–1482. doi:10.1016/j.acra.2015.08.015.

Recall rate reduction with tomosynthesis during baseline screening examinations – an assessment from a prospective trial

Jules H. Sumkin, DO[1], **Marie A. Ganott, MD**, **Denise M. Chough, MD**, **Victor J. Catullo, MD**, **Margarita L. Zuley, MD**, **Dilip D. Shinde, MD**, **Christiane M. Hakim, MD**, **Andriy I. Bandos, PhD**[2], and **David Gur, ScD**[3]

[1]Magee-Womens Hospital, Department of Radiology, Breast Imaging, 300 Halket Street, Pittsburgh, PA, 15213; Fax: (412)641-1414

[2]University of Pittsburgh, Graduate School of Public Health, Department of Biostatistics, A414 Crabtree Hall, 130 DeSoto Street, Pittsburgh, PA 15261; Fax: 412-624-0184

[3]University of Pittsburgh, School of Medicine, Department of Radiology, Radiology Imaging Research, 3362 Fifth Avenue, Pittsburgh, PA, 15213; Fax: (412)641-2582

Abstract

Rational and Objectives—Assess results of a prospective, single site clinical study evaluating digital breast tomosynthesis (DBT) during baseline screening mammography.

Materials and Methods—Under an institutional review board approved HIPAA compliant protocol, consenting women between ages 34 and 56 scheduled for their initial/baseline screening mammogram underwent both Full Field Digital Mammography (FFDM) and DBT. The FFDM and the FFDM plus DBT images were interpreted independently in a reader by mode balanced approach by two of 14 participating radiologists. A woman was recalled for a diagnostic workup if either radiologist recommended a recall. We report overall recall rates and related diagnostic outcome from the 1080 participants. Proportion of recommended recalls (BIRADS 0) were compared using a generalized linear mixed model (SAS 9.3) with a significance level of p=0.0294.

Corresponding author. gurd@upmc.edu, Phone: 412-641-2513.
Jules H. Sumkin, jsumkin@upmc.edu, Phone: 412-641-1976
Marie A. Ganott, mganott@magee.edu, Phone: 412-641-1409
Denise M. Chough, dchough@magee.edu, Phone: 412-641-1409
Victor J. Catullo, vcatullo@magee.edu, Phone: 412-641-1635
Margarita L. Zuley, zuleyml@upmc.edu, Phone: 412-641-4709
Dilip D. Shinde, dshinde@magee.edu, Phone: 412-641-1635
Christiane M. Hakim, chakim@magee.edu, Phone: 412-641-1965
Andriy I. Bandos, anb61@pitt.edu, Primary Phone: 412-383-5738

Disclosure: Jules Sumkin, Margarita Zuley, and Andriy Bandos are principal investigators on grants funded by Hologic, Inc. to the University of Pittsburgh. No disclosures for Marie Ganott, Denise Chough, Victor Catullo, Dilip Shinde, Christiane Hakim, and David Gur.

IRB statement: Approved by the Institutional Review Board of the University of Pittsburgh

Results—The fraction of women without breast cancer recommended for recall using FFDM alone and FFDM plus DBT were 412/1074 (38.4%) and 274/1074 (25.5%), respectively ($p<0.001$). Large inter-reader variability in terms of recall reduction was observed among the 14 readers; however, 11 out of 14 readers recalled fewer women using FFDM plus DBT (5 with p-values <0.015). Six cancers (4 DCIS and 2 IDC) were detected. One IDC was detected only on DBT and one DCIS cancer was detected only on FFDM, while the remaining cancers were detected on both modalities.

Conclusion—The use of FFDM plus DBT resulted in a significant decrease in recall rates during baseline screening mammography with no reduction in sensitivity.

Keywords

Mammography; Screening; Initial Examination; Recall; Tomosynthesis

INTRODUCTION

As wide spread periodic mammographic screening is now an acceptable practice in the United States and many other countries, our understanding of strategic, operational, and financial issues related to this practice is continually improving. Several performance measures have been used to define practice parameters in screening mammography, such as sensitivity, specificity, recall rate, positive predictive value (PPV), person-year-saved per examination, and cost per detected cancer [1, 2]. To date, despite the continuing controversy about the impact of recall rates on the overall cost benefit of screening mammography [3], the primary focus in screening has been on improving sensitivity. Although studies have shown that women who had false-positive mammograms remain likely to return for subsequent screening [4, 5], there is still some uncertainty regarding the possible effects false-positive mammograms may have on future compliance and participant attitudes toward screening [6]. This may especially be true for younger women participating in screening mammography for the first time for whom there are no prior images for comparison and who do not have previous experience with undergoing the procedure or being called back for further evaluation. As expected, higher recall rates than those during repeat screening have been reported in women with no prior mammograms [7–11]. This issue raises concern because, in addition to the operational and financial burden [12], women who have been recalled experience an added level of anxiety [13, 14]. There is a general belief that through a variety of actions, including possibly targeted training, radiologists' performance levels could be improved in this regard [15, 16]. However, to our knowledge, currently, there is no focused effort and/or specific training related to the interpretation of baseline mammograms (i.e., without priors).

Although not specifically regulated, there is a practice guideline in the United States to maintain an overall recall rate benchmark below 10% for the general mammography screening population that includes a mix of baseline and repeat screening [15, 17, 18]. The question of what effect, if any, does a forced reduction of recall rates have on detection rates remains somewhat controversial. However, there is a widely accepted belief that despite a demonstrated correlation between recall rates and cancer detection rates, it is important to keep recall rates as low as reasonably achievable [16,19]. One possible approach to reduce

recall rates in baseline screening mammography procedures is to use Digital Breast Tomosynthesis (DBT) as a recommended Standard of Practice [20]. DBT offers an approximation to a three-dimensional viewing of the breast, thereby eliminating some of the difficulties in correctly interpreting mammograms due to overlapping imaged tissue [21], and has been shown to reduce overall recall rates in retrospective studies and in clinical practice [22–30]. However, there are no reports to date focusing specifically on the use of DBT in baseline mammograms. We report here on a single institution prospective screening study that included independent viewing and interpretation of Full Field Digital Mammography (FFDM) alone versus FFDM plus DBT acquired on younger women receiving baseline examinations. The decision to focus on younger women receiving baseline screening was solely based on the assumption that this population could potentially benefit the most in terms of a reduction in recall rates when using DBT during a baseline screening examination.

MATERIALS AND METHODS

Study Population

Subject recruitment for this study began in May, 2010 and ended in September, 2014. Inclusion criteria was asymptomatic women between 34 and 56 years of age presenting for their baseline screening mammogram at our facility for any reason, including intentionally recruiting women enrolled in our high risk clinic because of our special interest in this group. Potential subjects were initially contacted by the research coordinator, if the potential subject gave her verbal permission documented by the technologist involved in the potential subject's clinical care. Each potential subject was given details of the study, along with an explanation of the tomosynthesis device, and written informed consent was obtained for each subject who agreed to participate. Women were excluded if they had a palpable finding by either self-examination or a clinical breast examination or if they were possibly pregnant or known pregnant by self-report. They were told that as participants in this research study they may be recalled at a higher than normal rate for additional imaging procedures (diagnostic workup). Approval of this HIPAA compliant study was obtained from the institutional review board at our institution.

Image Acquisition

Each consenting woman received an FFDM and DBT baseline examination that consisted of mediolateral oblique (MLO) and craniocaudal (CC) views for each breast. The FFDM and DBT images were acquired on a tomosynthesis system (Selenia Dimension, Hologic, Inc, Bedford, MA) that is capable of acquiring both FFDM and DBT images during a single breast compression. The image acquisitions took approximately the same amount of time within the total allotted examination time in our clinical practice. The system computed fibro-glandular radiation dose per view for the sets of 2D –FFDM projection mammograms and the corresponding DBT images included in the study. The average compressed breast thicknesses in this group of women were 56mm +/− 14mm and 58mm +/−17mm for CC and MLO views, respectively. For the FFDM procedures, average system computed fibro-glandular radiation dose levels per view were 1.96 +/− 0.72 mGy and 2.08+/−0. 74 mGy for the CC and MLO views, respectively. For the CC and MLO tomosynthesis procedures,

average system computed dose levels were 2.10+/−0.65mGy and 2.21+/−0.77mGy, respectively.

Image Interpretation

Images were transferred to two workstations for later interpretation. One workstation was uploaded only with the FFDM examination and the other with both FFDM and DBT examinations. All available features that are available and used during routine clinical practice, such as computer aided-detection (CAD) for the FFDM images, were also available to the interpreters in this prospective experiment. The workstations are physically located in separate rooms and the 14 participating radiologists were instructed NOT to discuss the case/exam with anyone until they rated the examination in question. This practice is not unique, since our facility performs over 70,000 screening procedures per year and it is quite rare that radiologists get the opportunity to, or deliberately make an attempt to, discuss specific screening cases before finalizing their interpretation and rating.

The FFDM images and the FFDM with DBT images were independently interpreted in the clinic by assigned experienced Mammography Quality Standards Act (MQSA) qualified and specifically trained radiologists, either on the same day of the examination or the following day. A total of 14 radiologists interpreted baseline examinations for this study with a range of 5 to 28 years of experience in interpreting screening mammography and 3 to 9 years of experience in interpreting FFDM plus DBT images. Participation of a radiologist depended on the clinical reading schedule; therefore, the number of examinations interpreted by each radiologist varied. However, the study coordinator independently assigned readings under each mode to different radiologists based on their scheduled assignments and availability in a manner designed to balance the number of examinations read by each radiologist under each of the two modes. Examination ratings were performed using the screening Breast Imaging Reporting and Data System (BIRADS) 0, 1, or 2, as BIRADS 3, 4, or 5 are not used in our screening operations. A subject was recalled for further diagnostic imaging workup if either of the interpreters (or both) recommended a recall (BIRADS 0), namely, by design we did not implement a consensus based decision step. Diagnostic imaging work-up for participants consisted of conventional care based on the recall recommendation of either (or both) radiologist. Any needed further evaluation (e.g. biopsy) was based on clinical practice standards.

Data analysis

The recommended recalls (BIRADS 0 ratings) from clinical interpretations of FFDM alone and FFDM combined with DBT were analyzed using generalized linear mixed model for binary data accounting for correlation between interpretations of examinations of the same patient and heterogeneity of the radiologists' performances under the considered modalities (proc glimmix, SAS v.9.3, Cary, NC). The primary statistical tests accounted for the fact that each of the examinations was interpreted by two different radiologists (one under each of the reading modes) and were performed at the two-sided significance level of 0.0294. The significance level was adjusted to account for a single interim analysis that had been reported previously [31] (also performed at 0.0294 level). The adjustment was originally planned using a generally utilized approach [32] to maintain an equal statistical test criterion

for the interim and final analyses and control the study's experiment-wise error rate at the 0.05 level. We also evaluated the number of benign biopsies eventually performed in the cohort of women initially recalled by either of the two modes. In addition, we assessed whether study conclusions would be affected by a breast based analysis (rather than case based), the exclusion of examinations performed on women younger than 40YO, and the total number of examinations interpreted by each of the radiologists. In particular, we excluded four radiologists (#1,#3,#5 and, #9), each of which interpreted only 16 examinations or fewer during the entire project, for assessing possible outliers.

RESULTS

The 1080 subjects who participated in this study were on average 42.03 years of age ± 3.75 years, including121 participants who were under the age of 40 at the time. Table 1 provides the distribution of the subjective breast density BIRADS ratings for women who participated in this study. As expected for this age range, the majority of participants were rated as having scattered and/or heterogeneously dense breast tissue.

The fraction of women without breast cancer recommended for recall using FFDM alone and FFDM combined with DBT was 412/1074 (38.4%) and 274/1074 (25.5%), respectively. This represents a 33% (138/412) reduction in recall rate ($p<0.001$). One hundred seventy six women without breast cancer (16.4%) were recalled by both modalities (table 2). Ten participating radiologists who interpreted more than 16 cases each reduced their recall rates, with five of these individuals (# 2, 4, 11, 12 and 13) reducing significantly their individual recall rate ($p<0.015$). Among the four radiologists who interpreted 16 or fewer examinations, two did not change their recall rate (#3 and #9), one decreased his/her recall rate (#1), and one who interpreted only three examinations (#5) while recalling one of two (50%) when using FFDM alone and one of one (100%) using FFDM plus DBT. The reduction observed in recall rates to date were quite variable among the 14 readers participating in the study. Table 3 provides the distribution of cases each radiologist interpreted and percentage of recalled cases for each mode. No significant trends were observed as the study progressed or since the interim analysis [31].

Table 4 summarizes the biopsies performed as a result of diagnostic workups that followed recalls. Of the 95 biopsies performed 34 benign biopsies were performed on women initially recalled by both modalities. Twenty-seven benign biopsies were performed on women recommended for a recall by the FFDM alone reader and 15 benign biopsies were performed on women recommended for a recall by the FFDM plus DBT reader only. Six biopsy proven cancer cases were found in the study population (4 ductal carcinoma in situ (DCIS) depicting only micro-calcifications and 2 invasive ductal carcinomas (IDC)) The average age of these subjects were 43.5 years, ± 2.59 years. One cancer was detected by the FFDM alone reading mode out of a total of 31 biopsies and one cancer was detected only by the FFDM plus DBT mode out of a total of 18 biopsies. Four cancers were detected in both modes out of 46 total biopsies. One DCIS case was recalled by the FFDM only reader and the FFDM plus DBT reader perceived the same DCIS case as benign, BIRADS 2 for micro-calcifications. The other DCIS cases were recalled by both readers under both modalities. Of

the two IDC cases, one of the cases was recalled by the FFDM+DBT interpreter only and was rated as an overall BIRADS 1 by the FFDM interpreter.

Study conclusions were not affected by a breast based analysis, the total number of examinations interpreted by each radiologist, or the exclusion of women who were younger than routine screening age (< 40YO).

DISCUSSION

The primary objective of this study was to evaluate recall rates of a single institution prospective clinical study involving independent viewing and interpretation of FFDM images alone versus FFDM+DBT acquired on younger women receiving baseline screening mammography, since their recall rates tend to be higher in general [7–11], and the recall rates in this group (women ≤ 50 years of age) was noted previously as particularly high in our own clinical practice (overall baseline recall rate of 28.4%) [33]; hence, this group is likely to benefit the most from a procedure that could potentially decrease recall rates. Indications from retrospective studies suggest that DBT in combination with FFDM should result in a significant decrease in recall rates when there are no prior examinations available for comparison [20, 22,24]. Prospective studies to date do not specify the impact of using DBT in baseline screening mammography and there are no published data addressing this group that disproportionately contributes to practice recall rates. Our results are in agreement in terms of the difference we observed. We note that during the study, our baseline recall rate using FFDM alone was quite similar to our overall clinical practice baseline recall rates. We reviewed all of our practice baseline recall rates for 2013 (>5600 examinations). The average recall rates were 37% and 27% for FFDM alone and FFDM +DBT, respectively. We suspect that our observations in terms of higher than expected recall rates for baseline mammograms are not unique to our institution and this is not a subgroup that is often evaluated carefully and adequately for practice parameters, in particular in stable practices where the fraction of baseline cases is quite small. In addition, we observed a substantial reduction in the performance of benign biopsies on women initially recalled under the FFDM plus DBT mode, despite the fact that once recalled the diagnostic workup was performed separately and often by a different radiologist than the originally interpreting one. We note that in this trial we did not measure time to interpretation under the two reading modes, as in this prospective study the interpretation of each of the cases was performed in the clinical environment and many of the interpretations were performed remotely from different sites.

The general concern regarding the approximate doubling of the radiation dose associated with using DBT in combination with FFDM, in particular for younger women, was not addressed in this project and may be eventually alleviated in the future through the use of synthesized 2D images [34, 35].

In this study of younger women we expected a few cancers to be detected and indeed six cancers were found during the study.

Our study has several limitations. First, we recognize the generalizability of the study results is limited, since the study is being conducted at one institution with a single group of radiologists on a specific group of women, and our clinical recall rate of baseline cases is high [33]. The low level of concordance in recalling women under the two modes highlights the large reader variability in interpreting baseline examinations, and, at least in our practice, the need to implement some type of intervention in this regard through one or more of several possible actions. This focused effort is planned in our future studies. For example, we intend to perform special second reader assessments of recommended "recalls" with feedback to the primary (original) interpreter in disagreement cases prior to finalizing his/her recommendations. Second, even though all women scheduled to undergo baseline on specific days were approached for possible consideration for participation in the study, we cannot rule out a self-selection bias. Third, by study design, we could have biased the outcomes in favor of more recalls on both modalities, because "study knowledge" tends to result in more aggressive ratings (i.e., additional recalls) [36, 37]. However, we decided for a patient to be recalled if either radiologist recommended a recall, since a consensus meeting for arbitrating disagreements, similar to that employed in some Europeans practices, could have biased the decision against FFDM. We do not believe that inclusion of women at high risk or the high underlying recall rate for conventional FFDM (2D) resulted in biasing the relative recall rate reduction we observed, albeit this could not be assessed in this study. Last, due to a small sample size combined with low expected prevalence in this age range, we could not assess quantitatively cancer detection rates of the two modalities being compared.

Conclusion

In this single institution prospective study we found that the use of FFDM plus DBT resulted in a significant decrease in recall rates during baseline screening mammography.

Acknowledgements

This work was supported in part by grant CA143019 from the National Cancer Institute, National Institutes of Health to the University of Pittsburgh

REFERENCES

1. Linver MN. Audits measure practice quality of mammography. Diagn Imaging. 2000; 22:57–61.
2. Burnside E, Belkora J, Esserman L. The impact of alternative practices on the cost and quality of mammographic screening in the United States. Clin Breast Cancer. 2001:145–152. [PubMed: 11899786]
3. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009; 151:716–726. [PubMed: 19920272]
4. Burman ML, Taplin SH, Herta DF, Elmore JG. Effect of false-positive mammograms on interval breast cancer screening in a health maintenance organization. Ann Intern Med. 1999 Jul 6; 131(1): 1–6. [PubMed: 10391809]
5. Pinckney RG, Geller BM, Burman M, Littenberg B. Effect of false-positive mammograms on return for subsequent screening mammography. Am J Med. 2003 Feb 1; 114(2):120–125. [PubMed: 12586231]
6. Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med. 2007 Apr 3; 146(7):502–510. Review. [PubMed: 17404352]

7. Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United kingdom. JAMA. 2003 Oct 22; 290(16):2129–2137. [PubMed: 14570948]
8. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. J Med Screen. 2005; 12(1):50–54. [PubMed: 15814020]
9. Yankaskas BC, Haneuse S, Kapp JM, Kerlikowske K, Geller B, Buist DS. Breast Cancer Surveillance Consortium. Performance of first mammography examination in women younger than 40 years. J Natl Cancer Inst. 2010 May 19; 102(10):692–701. [PubMed: 20439838]
10. Schell MJ, Yankaskas BC, Ballard-Barbash R, et al. Evidence-based target recall rates for screening mammography. Radiology. 2007 Jun; 243(3):681–689. [PubMed: 17517927]
11. Törnberg S, Kemetli L, Ascunce N, et al. A pooled analysis of interval cancer rates in six European countries. Eur J Cancer Prev. 2010 Mar; 19(2):87–93. [PubMed: 20010429]
12. Chubak J, Boudreau DM, Fishman PA, Elmore JG. Cost of breast-related care in the year following false positive screening mammograms. Med Care. 2010 Sep; 48(9):815–820. [PubMed: 20706161]
13. Brett J, Austoker J. Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. J Public Health Med. 2001; 23:292–300. [PubMed: 11873891]
14. Sandin B, Chorot P, Valiente RM, Lostao L, Santed MA. Adverse psychological effects in women attending a second-stage breast cancer screening. J Psychosom Res. 2002; 52:303–309. [PubMed: 12023127]
15. Feig SA. Economic challenges in breast imaging. A survivor's guide to success. Radiol Clin North Am. 2000; 38:843–852. [PubMed: 10943281]
16. Sickles EA. Successful methods to reduce false-positive mammography interpretations. Radiol Clin North Am. 2000; 38:693–700. [PubMed: 10943271]
17. U.S Department of Health and Human Services, Agency for Health Care Policy and Research. Clinical practice guideline number 13: Quality determinants of mammography; 1994 Oct. p. 78-86.AHCPR Publication No. 95-0632
18. Schell MJ, Yankaskas BC, Ballard-Barbash R, et al. Evidence-based target recall rates for screening mammography. Radiology. 2007; 243(3):681–689. [PubMed: 17517927]
19. Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of recall rates with sensitivity and positive predictive values of screening mammography. AJR. 2001; 177:543–549. [PubMed: 11517044]
20. FDA U.S. Food and Drug Administration. Medical Devices. Selenia Dimensions 3D System - P080003. http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm246400.htm
21. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. Radiology. 1997 Nov; 205(2):399–406. [PubMed: 9356620]
22. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. AJR Am J Roentgenol. 2009 Aug; 193(2):586–591. [PubMed: 19620460]
23. Gennaro G, Toledano A, di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. Eur Radiol. 2010; 20:1545–1553. [PubMed: 20033175]
24. Poplack SP, Tosteson TD, Kogel CA, et al. Digital breast tomosynthesis: Initial experience in 98 women with abnormal digital screening mammography. Am J Roentgenol. 2007; 189:616–623. [PubMed: 17715109]
25. Teertstra HJ, Loo CE, van den Bosch MAAJ, et al. Breast tomosynthesis in clinical practice: initial results. Eur Radiol. 2010; 20:16–24. [PubMed: 19657655]
26. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. J Natl Cancer Inst. 2014 Oct 13.106(11)

27. Lourenco AP, Barry-Brooks M, Baird G, Tuttle A, Mainiero MB. Changes in Recall Type and Patient Treatment Following Implementation of Screening Digital Breast Tomosynthesis. Radiology. 2015 Feb; 274(2):337–342. [PubMed: 25247407]

28. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. Radiology. 2015 Jan; 274(1):85–92. [PubMed: 25188431]

29. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA. 2014 Jun 25; 311(24):2499–2507. [PubMed: 25058084]

30. Skaane P1, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology. 2013 Apr; 267(1):47–56. [PubMed: 23297332]

31. BLINDED.

32. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977; 64:191–199.

33. BLINDED.

34. Gur D, Zuley ML, Anello MI, et al. Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. Acad Radiol. 2012 Feb; 19(2):166–171. [PubMed: 22098941]

35. Skaane P, Bandos AI, Eben EB, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. Radiology. 2014 Jun; 271(3):655–663. [PubMed: 24484063]

36. Ciatto S, Del Turco MR, Morrone D, et al. Independent double reading of screening mammograms. J Med Screen. 1995; 2(2):99–101. [PubMed: 7497164]

37. Ciatto S, Ambrogetti D, Bonardi R, et al. Second reading of screening mammograms increases cancer detection and recall rates. Results in the Florence screening programme. J Med Screen. 2005; 12(2):103–106. [PubMed: 15949122]

Table 1

Distribution of subjective breast density BIRADS ratings with average age in each of the density groups.

BIRADS†	Number of Cases	Percentage	Average Age	Standard deviation
1	54	5.0%	43.8	± 4.20
2	433	40.1%	42.1	± 3.82
3	549	50.8%	41.9	± 3.65
4	44	4.1%	40.7	± 2.90

† BIRADS: Breast Imaging-Reporting and Data System, 1= almost all fatty, 2= scattered, 3= heterogeneous, 4= extremely dense

Table 2

The number of cases recalled (and not recalled) by the different interpreters under each of the reading modes in women without breast cancer.

		FFDM Only		
		Recalled	**Not Recalled**	**Total**
FFDM+DBT	Recalled	176	98	274
	Not Recalled	236	564	800
	Total	412	662	1074

Table 3

Number of baseline screening interpretations by each of the readers and the number of examinations recommended for a recall by reader and mode in women without breast cancer.

Reader	FFDM			FFDM + DBT			
	Number of cases	Number of Recalls	(%)	Number of cases	Number of Recalls	(%)	p-value
Reader 1	4	1	(25)	4	0	(0)	1.00
Reader 2	114	35	(31)	127	18	(14)	< 0.01
Reader 3	8	4	(50)	8	4	(50)	1.00
Reader 4	156	33	(21)	175	19	(11)	0.01
Reader 5	2	1	(50)	1	1	(100)	1.00
Reader 6	51	18	(35)	63	21	(33)	0.85
Reader 7	175	63	(36)	158	46	(30)	0.20
Reader 8	179	91	(51)	180	73	(41)	0.06
Reader 9	1	1	(100)	1	1	(100)	1.00
Reader 10	51	17	(33)	53	15	(28)	0.67
Reader 11	53	33	(62)	54	19	(35)	<0.01
Reader 12	89	39	(44)	85	14	(16)	<0.01
Reader 13	140	56	(40)	129	30	(23)	<0.01
Reader 14	51	20	(39)	36	13	(38)	0.83
TOTALS	**1074**	**412**	(38)	**1074**	**274**	**(26)**	**<0.001**

Table 4

Biopsy outcome by mode of interpretation of the baseline screening examination that resulted in a recall recommendation.

	Biopsy Outcome			
Recalled By	**Benign**	**High Risk**	**Cancer**	**Totals**
FFDM only reader	27	3	1	**31**
FFDM+DBT reader	15	2	1	**18**
Both readers	34	8	4	**46**
Totals	**76**	**13**	**6**	**95**