Original Investigation

Recall Rate Reduction with Tomosynthesis During Baseline Screening Examinations:

An Assessment From a Prospective Trial

Jules H. Sumkin, DO, 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, David Gur, ScD

Rationale 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 Health Insurance Portability and Accountability Act (HIPAA)-compliant protocol, consenting women between ages 34 and 56 years scheduled for their initial and/or 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 work-up if either radiologist recommended a recall. We report overall recall rates and related diagnostic outcome from the 1080 participants. Proportion of recommended recalls (Breast Imaging Reporting and Data System 0) were compared using a generalized linear mixed model (SAS 9.3) with a significance level of $P = .0294$.

Results: The fraction of women without breast cancer recommended for recall using FFDM alone and FFDM plus DBT were 412 of 1074 (38.4%) and 274 of 1074 (25.5%), respectively ($P < .001$). Large inter-reader variability in terms of recall reduction was observed among the 14 readers; however, 11 of 14 readers recalled fewer women using FFDM plus DBT (5 with $P < .015$). Six cancers (four ductal carcinomas in situ [DCIS] and two invasive ductal carcinomas [IDC]) were detected. One IDC was detected only on DBT and one DCIS cancer was detected only on FFDM, whereas the remaining cancers were detected on both modalities.

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

Key Words: 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, 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

Acad Radiol 2015; ■:1–6

From the Department of Radiology, Magee-Womens Hospital, Pittsburgh, Pennsylvania (J.H.S., M.A.G., D.M.C., V.J.C., M.L.Z., D.D.S., C.M.H.); Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (A.I.B.); and Department of Radiology, Radiology Imaging Research, School of Medicine, University of Pittsburgh, 3362 Fifth Ave., Pittsburgh, PA 15213 (D.G.). Received July 7, 2015; accepted August 14, 2015. Disclosure: J.H.S., M.L.Z., and A.I.B. are principal investigators on grants funded by Hologic, Inc. to the University of Pittsburgh. No disclosures for other authors. IRB statement: Approved by the Institutional Review Board of the University of Pittsburgh. **Address correspondence to:** D.G. e-mail: gurd@upmc.edu

http://dx.doi.org/10.1016/j.acra.2015.08.015

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 (ie, 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 (10,15,17). 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 (16,18). 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 (19). DBT offers an approximation to a three-dimensional viewing of the breast, thereby eliminating some of the difficulties in correctly interpreting mammograms because of overlapping imaged tissue (20), and has been shown to reduce overall recall rates in retrospective studies and in clinical practice (21–29). 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 aged between 34 and 56 years 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 work-up). 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 fibroglandular radiation dose per view for the sets of two-dimensional (2D)–FFDM projection mammograms and the corresponding DBT images included in the study. The average compressed breast thicknesses in this group of women were 56 ± 14 mm and 58 ± 17 mm for CC and MLO views, respectively. For the FFDM procedures, average system-computed fibroglandular 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 ± 0.65 mGy and 2.21 ± 0.77 mGy, 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 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/examination with anyone until they rated the examination in question. This practice is not unique because 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–qualified and specifically trained radiologists, either on the same day of the examination or the following day. A total of 14 radiologists