Comparison of Digital Mammography Alone and Digital Mammography Plus Tomosynthesis in a Population-based Screening Program[1]

Per Skaane, MD, PhD
Andriy I. Bandos, PhD
Randi Gullien, RT
Ellen B. Eben, MD
Ulrika Ekseth, MD
Unni Haakenaasen, MD
Mina Izadi, MD
Ingvild N. Jebsen, MD
Gunnar Jahr, MD
Mona Krager, MD
Loren T. Niklason, PhD
Solveig Hofvind, PhD
David Gur, ScD

Purpose: To assess cancer detection rates, false-positive rates before arbitration, positive predictive values for women recalled after arbitration, and the type of cancers detected with use of digital mammography alone and combined with tomosynthesis in a large prospective screening trial.

Materials and Methods: A prospective, reader- and modality-balanced screening study of participants undergoing combined mammography plus tomosynthesis, the results of which were read independently by four different radiologists, is under way. The study was approved by a regional ethics committee, and all participants provided written informed consent. The authors performed a preplanned interim analysis of results from 12631 examinations interpreted by using mammography alone and mammography plus tomosynthesis from November 22, 2010, to December 31, 2011. Analyses were based on marginal log-linear models for binary data, accounting for correlated interpretations and adjusting for reader-specific performance levels by using a two-sided significance level of .0294.

Results: Detection rates, including those for invasive and in situ cancers, were 6.1 per 1000 examinations for mammography alone and 8.0 per 1000 examinations for mammography plus tomosynthesis (27% increase, adjusted for reader; $P = .001$). False-positive rates before arbitration were 61.1 per 1000 examinations with mammography alone and 53.1 per 1000 examinations with mammography plus tomosynthesis (15% decrease, adjusted for reader; $P < .001$). After arbitration, positive predictive values for recalled patients with cancers verified later were comparable (29.1% and 28.5%, respectively, with mammography alone and mammography plus tomosynthesis; $P = .72$). Twenty-five additional invasive cancers were detected with mammography plus tomosynthesis (40% increase, adjusted for reader; $P < .001$). The mean interpretation time was 45 seconds for mammography alone and 91 seconds for mammography plus tomosynthesis ($P < .001$).

Conclusion: The use of mammography plus tomosynthesis in a screening environment resulted in a significantly higher cancer detection rate and enabled the detection of more invasive cancers.

Clinical trial registration no. NCT01248546

[1] From the Department of Radiology, Breast Imaging Center, Oslo University Hospital Ullevaal, University of Oslo, Kirkeveien 166, N-0407 Oslo, Norway (P.S.); Departments of Biostatistics (A.I.B.) and Radiology (D.G.), University of Pittsburgh, Pittsburgh, Pa; Department of Radiology, Oslo University Hospital, Oslo, Norway (R.G., E.B.E., U.H., M.I., I.N.J., G.J., M.K.); Curato Roentgen Institute, Oslo, Norway (U.E.); Hologic, Bedford, Mass (L.T.N.); and Institute of Population-based Cancer Research, The Cancer Registry, Oslo, Norway (S.H.). Received June 20, 2012; revision requested July 31; revision received August 20; accepted September 7; final version accepted September 19. **Address correspondence to** P.S. (e-mail: *perska@ous-hf.no*).

Mammographic screening for the early detection of breast cancer reduces mortality and is a widely accepted practice in many countries (1,2). Despite recent controversy regarding the benefits of mammographic screening, mammography remains the most commonly used procedure for this purpose (3,4). Because of the recognized limitations of mammography, in particular the issue of overlapping imaged tissue, a number of x-ray– and non-x-ray–based procedures are being investigated as possible replacements for or adjunct modalities to mammography, whether for universal use or for use in specific subpopulations (5–8). Each of these approaches represents a separate procedure that necessitates substantial additional technical and professional resources.

The ability to perform tomosynthesis reconstruction from limited two-dimensional data sets has been known, but it was only with the advent of large-area, fast-reading digital detectors that it became possible to incorporate these technologies into routine clinical practice (9–12). The feasibility of using tomosynthesis in breast imaging was demonstrated more than a decade ago (13), and the U.S. Food and Drug Administration recently approved the first commercial system for clinical use (14). Digital mammography–based tomosynthesis produces cross-sectional sections by using multiple, low-dose acquisitions with total radiation exposure and breast compression similar to that used in conventional mammography. A number of small retrospective studies in which investigators evaluated tomosynthesis primarily in laboratory settings by using cancer-enriched populations demonstrated the potential for decreasing recall rates and possibly increasing cancer detection rates; however, none of these studies was performed in a manner that could affect treatment decisions (15–22). We performed this study to assess cancer detection rates, false-positive rates before arbitration, positive predictive values for women recalled after arbitration, and the type of cancers detected with use of digital mammography alone and combined with tomosynthesis in a large prospective screening trial. Because of the complexities of the four-arm Oslo Tomosynthesis Screening Trial and because two of these arms were designed specifically to assess the possible effects of different experimental parameters on performance, we focus herein on a preplanned interim analysis of two arms—one constituting the commonly used practice of mammography alone and one constituting a mammography plus tomosynthesis arm approved by the U.S. Food and Drug Administration that operationally requires only minor modifications to the conventional examination.

Advances in Knowledge

- The overall performance level during the interpretation of screening mammograms improved significantly with the addition of tomosynthesis; the combination of tomosynthesis and digital mammography resulted in a significantly higher cancer detection rate (27% increase, $P = .001$) and a reduction in false-positive findings (15%, $P < .001$) compared with digital mammography alone.
- A significantly higher detection rate for invasive cancers (40% increase, $P < .001$) was observed for the interpretation of screening mammograms with tomosynthesis.
- Interpretation time was longer for digital mammography combined with tomosynthesis than for digital mammography alone (91 vs 45 seconds, respectively; $P < .001$).

Implication for Patient Care

- The implementation of tomosynthesis in breast cancer screening will improve cancer detection, especially for invasive cancer, and should also reduce an individual radiologist's recall rate.

Materials and Methods

Hologic (Bedford, Mass) sponsored the study by providing tomosynthesis equipment and financial support for additional interpretations. L.T.N. and D.G. are employed by Hologic. The Department of Biostatistics at the University of Pittsburgh, Penn (A.I.B.), was contracted to perform all statistical analyses independently. D.G. performed the analyses for Hologic at the time that revisions were being processed. The remaining authors were not employed by Hologic and were without potential conflicts of interest; these authors had control of the data and presented the material as submitted for publication.

Study Group

This large, single-institution prospective study was approved by a regional ethics committee, and written informed consent was obtained from all participants. Herein, we summarize results of a preplanned interim analysis of two arms of a four-arm study covering all consenting participants during the period from November 22, 2010, to December 31, 2011. Figure 1 shows a detailed study

Published online before print
10.1148/radiol.12121373 **Content code:** BR

Radiology 2013; 267:47–56

Abbreviations:
CI = confidence interval
PPV = positive predictive value

Author contributions:
Guarantors of integrity of entire study, P.S.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, P.S., R.G., E.B.E., U.E., U.H., M.I., I.N.J., G.J., M.K., L.T.N., S.H., D.G.; clinical studies, P.S., R.G., E.B.E., U.E., U.H., M.I., I.N.J., G.J., M.K.; statistical analysis, P.S., A.I.B., S.H.; and manuscript editing, P.S., R.G., E.B.E., U.E., U.H., M.I., I.N.J., G.J., M.K., L.T.N., S.H., D.G.

Clinical trial registration no. NCT01248546

Conflicts of interest are listed at the end of this article.

flowchart. A total of 29652 women (age range, 50–69 years) were invited by means of personal letter to undergo routine, biennial, two-view (craniocaudal and mediolateral oblique) screening mammography during this period as part of the Oslo screening program. Of the 29652 women, 17960 (60.6%) attended the screening program, with 12631 consenting to participate in the study. The Oslo screening program has been described in detail previously (23,24). The program is part of the Norwegian Breast Cancer Screening Program administered by the Cancer Registry of Norway. Independent double reading was the standard of practice, and mammograms considered suspicious for malignancy by at least one reader were discussed at a consensus-based arbitration meeting. Diagnostic work-up, including additional views, ultrasonography, magnetic resonance imaging, and needle biopsy (if indicated), was performed by the same group of radiologists. Short-term follow-up imaging was not used. Disabled women (eg, unable to stand) and women with breast implants were excluded.

During the current screening cycle, a selected group of women were asked to participate in the study. Potential candidates were selected on the basis of whether technical staff members and imaging systems were available to perform the additional imaging examination and not on the basis of any personal information about the women. Before commencement, all radiographers, radiologists, and staff were trained specifically in how to obtain and interpret tomosynthesis images. All readers received individualized intensive personal training (lasting approximately 4 hours) in reviewing examination results by using the same workstation and hanging protocols. In addition, seven readers (P.S., E.B.E., U.E., U.H., I.N.J., G.J., M.K.) had previously taken part in a pilot experimental clinical study of tomosynthesis (25), and one reader (M.I.) underwent the same training program used in the pilot study. As a result, training of each participating radiologist included a detailed review with feedback of enriched sets with a minimum of 100 examinations before participation as a reader in the prospective trial.

Figure 1

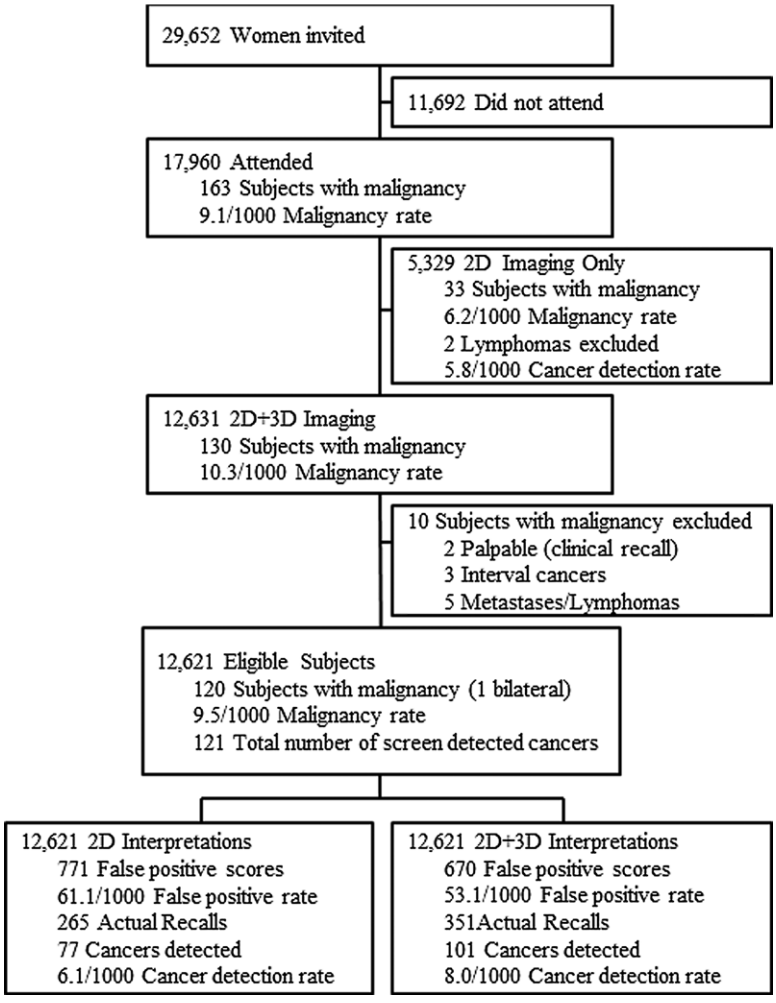

Figure 1: Study flowchart. Malignancy rate is number of subjects with malignancy per 1000 women screened. Cancer detection rate is based on breast-based scoring and includes only screening-detected primary breast cancers. After exclusion of 10 patients with nonscreening-detected breast malignancies ($n = 5$) and nonbreast primary cancers ($n = 5$), 121 cancers were detected in 120 women. The 121 cancers may have been detected by any of the four radiologists reading cases in the study. Of these 121 cancers, 77 were detected with mammography alone (cancer detection rate, 6.1 of 1000 women) and 101 were detected with mammography plus tomosynthesis (cancer detection rate, 8.0 of 1000 women). *2D* = mammography alone, *2D+3D* = mammography plus tomosynthesis.

Imaging Techniques

Two views (craniocaudal and mediolateral oblique) were obtained of each breast with digital mammography and tomosynthesis by using a commercially available system (Dimensions, Hologic), with single breast compression per view. The additional time required to obtain the tomosynthesis images was about 10 seconds per view. The radiation dose levels for mammography plus tomosynthesis combined were approximately twice those for mammography alone. All images were transferred to the Breast Imaging Center at Ullevaal University Hospital for interpretation and treatment recommendation.

Table 1

Scale Used to Score Breasts and Cases

Score	Description
1	Negative for suggestive findings or definitely benign
2	The breast (case) would have been recalled for a probably benign finding, but the interpreter referred final decision to an arbitration meeting
3	Recalled by the interpreter for a suggestive finding, but he or she wants the case discussed with the final decision made at the arbitration meeting
4	Recalled by the interpreter for a probably malignant finding, and the woman must be recalled for additional imaging and needle biopsy; not allowed to dismiss at arbitration meeting
5	Recalled by the interpreter for a highly suggestive finding of typical malignancy; a benign diagnosis at needle biopsy is not expected, so surgical biopsy is mandatory if the needle biopsy result is negative or inconclusive

Image Interpretation

Images from examinations performed with each modality were interpreted independently in a batch mode by using a dedicated workstation. Eight radiologists with 2–31 years of experience in screening mammography participated in this study. Images were interpreted independently by four radiologists by using each of the reading modes. A scheduler determined independently the specific cases to be interpreted and which of the eight radiologists would be assigned (by session and mode) to interpret them. The scheduler attempted to balance the number of cases interpreted by each radiologist with each modality. The four arms included mammography alone, mammography plus computer-aided detection, mammography plus tomosynthesis, and synthesized mammography plus tomosynthesis (in which mammographic images were reconstructed from the three-dimensional data set). Hanging protocols for the interpreter's assigned arm were preset; hence, radiologists had access only to the images required for the assigned arm. If available, previous screening mammograms were reviewed. Hanging protocols for mammography plus tomosynthesis initially displayed the mammograms alone (similar to the mammography hanging protocol) and then displayed the mammograms for each view on the left monitor and the tomosynthesis image for the same view on the right monitor.

Each radiologist independently rated the images from each examination by using astandardized five-point rating scale (Table 1). Furthermore, if a reader recorded a positive score (≥2), mammographic features had to be specified. Scores were recorded directly into the Norwegian Breast Cancer Screening Program database and locked after each reading session. Interpretation time was recorded automatically.

Arbitration Meeting

All cases that received at least one score of 2 or greater in at least one arm were discussed at arbitration before a consensus-based clinical treatment decision was made. All imaging and nonimaging information was made available. A minimum of two screening radiologists participated in these consensus-based arbitration meetings, during which a binary decision was made to *(a)* dismiss the initially suspected findings or *(b)* invite the participant to return for diagnostic work-up. A consensus-based breast parenchyma density score was assigned according to the American College of Radiology Breast Imaging Reporting and Data System (categories 1–4).

Findings Categories and Summary Measures

The primary end points of this prospective screening study are based on comparisons of cancer detection rates, false-positive rates before arbitration, and diagnostic performance in terms of positive predictive value (PPV) for patients recalled after arbitration. A cancer confirmed as a result of arbitration and follow-up diagnostic work-up for which the screening score was 2 or higher with use of a specific modality was considered detected with that modality and attributed to that modality as a true-positive finding. Conversely, a participant without a verified cancer who originally was referred to arbitration according to results of a specific modality but who was later dismissed at arbitration was attributed to that modality as a false-positive event. These definitions of true-positive and false-positive events are analogous to paradigms used in other detection and localization studies (26,27). Cancer detection was verified for participants referred to arbitration according to the results of at least one of the screening modalities, and only limited data about interval cancers are available at this time; therefore, we cannot estimate conventional absolute sensitivity or specificity. However, because of the paired design of the study, namely that the technologies being compared are applied simultaneously to the same population, we are able to estimate relative performance levels rather than absolute sensitivity or 1 − specificity (28–31).

We defined the cancer detection rate attained with each modality as the number of detected cancers attributed to that modality (ie, screening score of 2 or higher) per 1000 screenings. The false-positive rate before arbitration for each modality was defined as the number of women per 1000 participants who were assigned a score of 2 or higher for that modality but were not found to have cancer. PPV was defined as the percentage of women who had received a score of 2 or higher during screening, were later recalled as a result of arbitration, and were found to have cancer (number of cancers divided by number of recalls).

Statistical Analyses

Because the technologies used in this trial continue to be developed, the interim analysis was planned primarily to assess whether all performance indicators were in accordance with the European Guidelines (32), whether we should fix all four arms in the study for

the duration, or whether we have sufficient reference data to consider adjusting one or more of the modalities at the end of phase I. As of January 2012, one of the experimental modalities (ie, the use of synthesized images) was changed substantially for phase II of the trial. We performed this interim analysis at a significance level of .0294 to maintain for the study an overall type I error rate at .05 while maintaining an equal significance criterion at the interim and final analyses (33). Inferences for ratios of the rates adjusted for differences among radiologists were conducted by using a Wald test in the context of a log-linear binary regression model (Proc Genmod, SAS 9.2; SAS Institute, Cary, NC), accounting for the correlation between interpretations related to examinations in the same patient (31). Confidence intervals (CIs) were obtained by means of exponentiation of the CIs around the estimated logarithms of these ratios.

Separate statistical models were built for comparisons of cancer detection and false-positive rates. Half the overall significance level (α = .0147 or .0294/2) was used in each model to enable simultaneous inferences. Both models were fit by using the entire screened cohort (17960 women); however, the comparison of rates was based on the mammography-plus-tomosynthesis subcohort (12621 women), and the mammography-alone subcohort contributed toward estimation of reader effects. The results from the full-cohort models were verified by using models fitted by using mammography-plus-tomosynthesis participants only. A similar analysis was performed for the detection rates for invasive cancers.

In the secondary analyses, we compared PPVs in women recalled for diagnostic work-up as a result of arbitration. Inferences for the odds ratio were conducted by using the Wald test, and the CIs were estimated in the context of logistic regression for correlated data. We also evaluated the consistency of our results after excluding seven cases of cancer referred for the wrong breast that were included in the primary analysis and after including 10 cases of known interval (n = 3), metastatic (n = 5, one bilateral), and palpable (n = 2) cancers that were excluded from the primary analysis.

Results

A total of 12631 women were enrolled in the study and underwent mammography plus tomosynthesis (Fig 1). After exclusion of 10 malignant cases, the remaining 12621 cases, including 121 screening-detected cancers, were included in the primary analysis.

Table 2 summarizes the cancer detection and false-positive rates as validated up to March 31, 2012, for each radiologist with mammography alone and with mammography plus tomosynthesis. The cancer detection rate was 6.1 per 1000 screenings (77 of 12621 cases) with mammography alone and 8.0 per 1000 screenings (101 of 12621 cases) with mammography plus tomosynthesis (31% increase). Because the images in different cases were read by each reader with use of each modality, prevalence varied within the groups. Hence, we provide the fraction of detected cancers in addition to detection rates. For seven of the eight radiologists, the fraction of cancers detected was increased with the addition of tomosynthesis. A large number of cancers were detected with mammography plus tomosynthesis by radiologist 5, who happened to be assigned to interpret more cases with cancer with this modality. After adjusting for differences among reader-specific performance levels, the ratio of cancer detection rates for mammography alone versus mammography plus tomosynthesis was 1.27 (P = .001; 98.5% CI: 1.06, 1.53).

The false-positive rate was 61.1 per 1000 screenings (771 of 12621 cases) with mammography alone and 53.1 per 1000 screenings (670 of 12621 cases) with mammography plus tomosynthesis (13% decrease) (Table 2). Five of the eight radiologists referred proportionally more patients for arbitration with use of mammography alone than with use of mammography plus tomosynthesis. After adjusting for differences among the reader-specific performance levels, the ratio of false-positive rates for mammography alone versus mammography plus tomosynthesis was 0.85 (P < .001; 98.5% CI: 0.76, 0.96). Overall, with 97% confidence (corresponding to a significance level of .0294), mammography plus tomosynthesis led to a simultaneous increase in cancer detection rates and decrease in false-positive rates.

Seventy-seven cancers were detected with mammography alone and 101 were detected with mammography plus tomosynthesis (a difference of 24 cancers, 31% increase) (Table 3). Notably, 25 additional invasive cancers (81 vs 56 cancers) were detected with mammography plus tomosynthesis (Fig 2). The corresponding reader-adjusted ratio of detection rates of 1.40 was significant (P < .001; 98.5% CI: 1.13, 1.71). The additional abnormal findings detected with mammography plus tomosynthesis were neither high-risk lesions nor ductal carcinoma in situ alone but rather were invasive cancers at detection. Although many of the cancers were primarily small and of lower grades at detection, 10 of the 25 additional cancers detected only with use of mammography plus tomosynthesis (40%) were grade 2 or higher (six cancers with grade 2 and four with grade 3), constituting a 26% (48 vs 38 cases) increase in the detection of higher-grade cancers (Table 3).

The PPVs in women who were recalled as a result of arbitration, which reconsidered all available imaging and other information, were similar for those initially assigned a positive score with mammography alone (77 women with cancer of 265 women with positive scores at mammography alone [29.1%]) and with mammography plus tomosynthesis (100 women with cancer, including one bilateral, of 351 women with positive scores at mammography plus tomosynthesis [28.5%]; P = .72; 95% CI for odds ratio: 0.72, 1.60). These results suggest that despite the fact that the number of cases assigned a positive score before arbitration and recalled after arbitration was greater with mammography plus tomosynthesis, there was no substantial decrease in the PPV

Table 2

Cancer Detection and False-Positive Rates according to Individual Radiologists

		Mammography Alone					Mammography Plus Tomosynthesis				
Radiologist	Years of Experience*	False-Positive Rate†	Cancer Detection Rate†	No. of Known Cancers‡	Detected Cancers (%)§	No. of Patients	False-Positive Rate†	Cancer Detection Rate†	No. of Known Cancers‡	Detected Cancers (%)§	No. of Patients
1	8	110.7 (80)	6.9 (5)	6	83.3	723	73.6 (46)	11.2 (7)	8	87.5	625
2	21	62.2 (175)	4.6 (13)	24	54.2	2812	68.3 (119)	7.5 (13)	15	86.7	1743
3	2	83.3 (131)	4.5 (7)	12	58.3	1573	55.3 (82)	4.7 (7)	9	77.8	1483
4	31	39.5 (64)	11.1 (18)	24	75.0	1622	44.4 (78)	5.1 (9)	10	90.0	1758
5	29	45.2 (106)	4.7 (11)	19	57.9	2346	52.7 (147)	13.3 (37)	43	86.0	2790
6	10	53.8 (78)	6.2 (9)	15	60.0	1451	50.6 (71)	5.7 (8)	14	57.1	1402
7	20	71.8 (67)	4.3 (4)	5	80.0	933	52.4 (71)	8.9 (12)	14	85.7	1355
8	6	60.3 (70)	8.6 (10)	16	62.5	1161	38.2 (56)	5.5 (8)	8	100.0	1465
All		61.1 (771)	6.1 (77)	121	63.6	12621	53.1 (670)	8.0 (101)	121	83.5	12621

* Years of experience in interpreting screening mammograms.

† False-positive and cancer detection rates are given as number per 1000 patients. Numbers in parentheses are numbers of patients.

‡ Known screening-detected cancer cases were those given a positive score of 2 or higher that were later proved to be cancer. The positive score may have come from any of the four radiologists who could refer cases to consensus (includes readers at mammography alone, mammography plus computer-aided detection, mammography plus tomosynthesis, and synthetic mammography plus tomosynthesis).

§ Percentage detected represents the fraction of known screening-detected cancers that were detected by the reader using the specific modality. For example, of the 121 known cancer cases, 77 were detected by all readers at mammography alone (77/121 = 63.6%).

in these patients. For the recalled cases originally assigned a positive score at either mammography alone or mammography plus tomosynthesis (but not both), the PPV was substantially higher for the cases initially referred at mammography plus tomosynthesis (27 of 167 cases [16.2%] vs five of 79 cases [6%]).

Our secondary analyses excluding the seven cancer cases with positive scores assigned to arbitration for the wrong breast and after including the 10 cancers excluded from the primary analysis indicated that the study results would not be meaningfully changed and all conclusions (statistical inferences) remained the same. We note that during an average of 9 months of follow-up to date we are aware of three interval cancers, but none changes the study conclusions because all three were assigned the same score with both mammography alone and mammography plus tomosynthesis.

As expected, most cases were rated as Breast Imaging Reporting and Data System density category 2 or 3. Notably, the additional cancers detected with mammography plus tomosynthesis were distributed across all breast densities, including fatty breasts (Table 3).

The mean (±standard deviation) compressed breast thickness during mammography plus tomosynthesis was 53.9 mm ± 12.8. The mean glandular dose during the mammography and the tomosynthesis imaging procedures were 1.58 mGy ± 0.61 and 1.95 mGy ± 0.58, respectively. The mean interpretation time was 45 seconds for mammography alone and 91 seconds for mammography plus tomosynthesis ($P < .001$).

Discussion

Tomosynthesis studies to date have focused primarily on the potential benefit of reducing false-positive interpretations (14,17) and were performed in experimental settings. In this large-scale prospective study, we investigated the potential benefits and limitations of tomosynthesis, if any, in a population-based mammographic screening

Table 3

Comparison of Cancers Detected at Mammography Alone and Combined with Tomosynthesis

Parameter*	All†	No. of Cancers Missed with Both Modalities‡	No. of Cancers Detected with 2D Only	No. of Cancers Detected with 2D+3D Only	No. of Cancers Detected with 2D and 2D+3D	Total with 2D	Total with 2D+3D	Difference
Total no. of cancers	121	14	6	30	71	77	101	24
Invasive cancers with or without DCIS	96	11	4	29	52	56	81	25
Histologic findings								
IDC	57	6	2	16	33	35	49	14
IDC and DCIS	19	3	0	5	11	11	16	5
ILC	17	2	2	7	6	8	13	5
Other primary invasive cancer	3	0	0	1	2	2	3	1
Grade								
1	37	4	1	16	16	17	32	15
2	44	6	3	9	26	29	35	6
3	13	0	0	4	9	9	13	4
Unknown	2	1	0	0	1	1	1	0
Lymph node status								
Negative	76	9	4	23	40	44	63	19
Positive	15	2	0	4	9	9	13	4
Unknown	5	0	0	2	3	3	5	2
Radiologic finding								
Circumscribed mass	9	0	0	2	7	7	9	2
Spiculated mass	43	3	3	12	25	28	37	9
Architectural distortion	20	4	0	8	8	8	16	8
Asymmetric density	6	1	1	1	3	4	4	0
Calcifications	6	0	0	0	6	6	6	0
Mass and calcifications	12	3	0	6	3	3	9	6
Breast density score								
1 (Fatty)	6	0	0	2	4	4	6	2
2 (Scattered)	44	7	1	11	25	26	36	10
3 (Heterogeneous)	40	3	3	14	20	23	34	11
4 (Extreme)	6	1	0	2	3	3	5	2
Lesion size (mm)								
≤10	45	6	3	12	24	27	36	9
11–15	27	4	0	13	10	10	23	13
16–19	6	0	1	0	5	6	5	−1
≥20	15	0	0	3	12	12	15	3
NA	3	1	0	1	1	1	2	1
Mean§	12.8	10.7	11.0	12.8	13.4	13.2	13.2	0
Median§	11	10	8.5	13	11	11	13	2
Range§	1–50	7–15	8–19	5–50	1–27	1–27	1–50	0–23
In situ cancers (DCIS)	25	3	2	1	19	21	20	−1
Grade								
Low or medium	5	1	0	0	4	4	4	0
High	20	2	2	1	15	17	16	−1
Radiologic sign								
Calcifications	23	2	2	1	18	20	19	−1
Mass and calcifications	2	1	0	0	1	1	1	0
Breast density score								
1 (Fatty)	2	0	0	0	2	2	2	0
2 (Scattered)	10	1	1	0	8	9	8	−1
3 (Heterogeneous)	13	2	1	1	9	10	10	0

Table 3 (continues)

Table 3 (continued)

Comparison of Cancers Detected at Mammography Alone and Combined with Tomosynthesis

Parameter*	All†	No. of Cancers Missed with Both Modalities‡	No. of Cancers Detected with 2D Only	No. of Cancers Detected with 2D+3D Only	No. of Cancers Detected with 2D and 2D+3D	Total with 2D	Total with 2D+3D	Difference
4 (Extreme)	0	0	0	0	0	0	0	0
Lesion size (mm)								
≤10	5	2	1	0	2	3	2	−1
11–15	8	0	1	0	7	8	7	−1
16–19	2	1	0	0	1	1	1	0
≥20	10	0	0	1	9	9	10	1
Mean§	22.0	8.0	10.0	85.0	22.2	21.0	25.3	4.3
Median§	15	3	10	85	17	15	18.5	3.5
Range§	3–85	3–18	9–11	85–85	5–50	5–50	5–85	0–35

Note.—Data are from 12 621 participants. 2D = mammography alone, 2D+3D = mammography plus tomosynthesis.

* DCIS = ductal carcinoma in situ, IDC = invasive ductal carcinoma, ILC = invasive lobular carcinoma, NA = not available. Ductal carcinoma in situ was graded according to the van Nuys classification.

† Includes all known cancers detected at screening by any of the four radiologists (includes readers at mammography alone, mammography plus computer-aided detection, mammography plus tomosynthesis, and synthetic mammography plus tomosynthesis).

‡ Cases missed by the readers at both mammography alone and mammography plus tomosynthesis that were detected at screening by one of the other two readers (at mammography plus computer-aided detection or synthetic mammography plus tomosynthesis).

§ Numbers are lesion sizes (in millimeters).

Figure 2

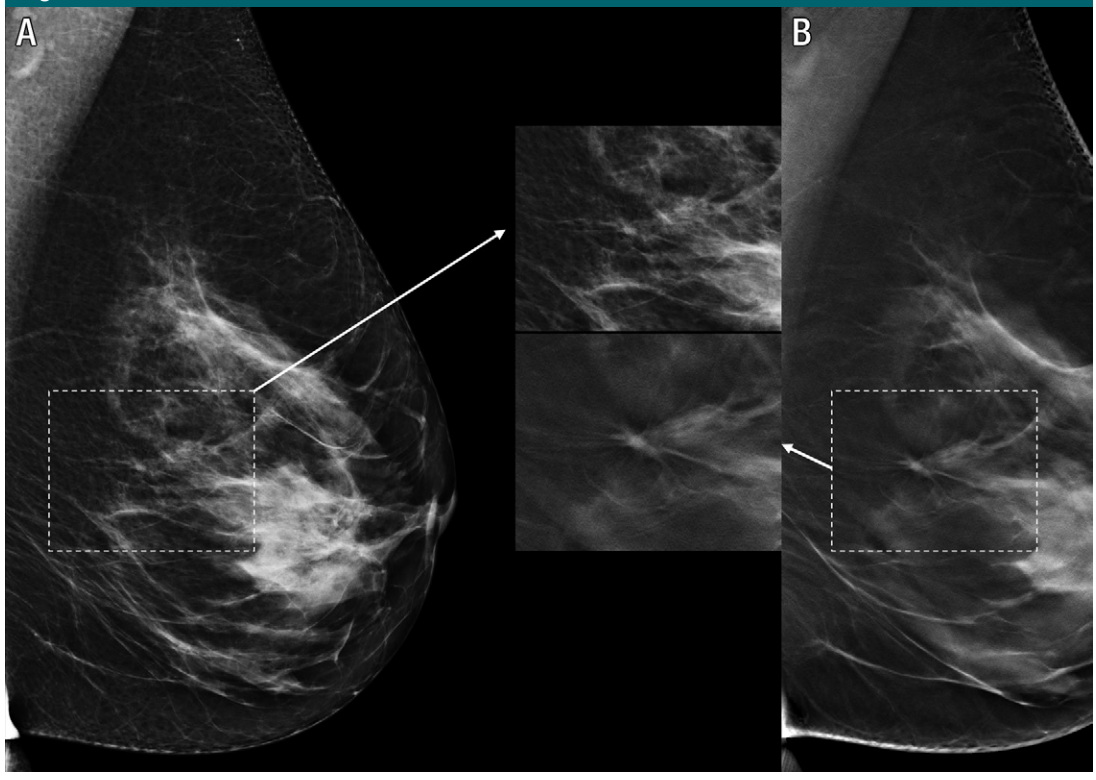

Figure 2: Mediolateral oblique views of left breast in 57-year-old woman recalled after mammographic screening because of a spiculated mass seen only at mammography plus tomosynthesis. *A,* Mammogram shows normal findings (score, 1). *B,* Tomosynthesis image demonstrates spiculated mass (score, 3). An 8-mm invasive ductal carcinoma was diagnosed at histologic examination.

program. Our results demonstrated that a substantial number of additional cancers were detected with use of mammography plus tomosynthesis versus mammography alone. We did not observe an improvement in the detection of ductal carcinoma in situ.

In addition, there were significantly fewer false-positive findings before arbitration with mammography plus tomosynthesis, albeit the reduction was smaller than suggested in a retrospective study (17). The overall actual number of women recalled as a result of arbitration was larger for those initially assigned a positive score at mammography plus tomosynthesis (351 vs 265 women). However, the concordant increase in the detection of 24 additional cancers resulted in a similar PPV for the cases ultimately recalled for arbitration. Because of the paired nature of the current study, these results could be biased against mammography plus tomosynthesis in that some of the dismissed cases initially referred on the basis of mammography alone might not have been dismissed if tomosynthesis had not been available at arbitration. Hence, the actual PPV based on mammography alone could be lower.

As related to possible overdiagnosis, in the context of breast cancer screening where the survival rate is high, it is virtually impossible to assess overdiagnosis in studies of less than 10–20 years duration. This issue remains controversial and is clearly beyond the scope of this interim analysis (34–36). With rapidly advancing technology and short product life cycles, one can only assess detection rates and focus on certain types of cancers one assumes that, left alone, would have high likelihood of eventually leading to mortality in at least some of the patients identified as having these types of cancers. We believe that many of the additional abnormal findings detected with mammography plus tomosynthesis in our study are the very types of cancers one would hope to detect early and treat. Perhaps our most important observation is that with the mammography-plus-tomosynthesis arm, the actual benefit, in terms of possibly improving outcome owing to earlier detection, may be larger than merely the difference in the total count or number of detected cancers.

As noted by others (37), the interpretation time was longer for mammography plus tomosynthesis than for mammography alone. However, optimization of interpretation efficiency by means of improved hanging protocols or work flow is beyond the scope of this investigation.

The current mammography-plus-tomosynthesis imaging procedure requires a radiation dose that is approximately double that for mammography alone. However, the radiation dose level for the combined examination was set to be below limits approved by the U.S. Food and Drug Administration, which constitutes an acceptable risk. In our study, with an average breast thickness of 54 mm, computed radiation dose levels for mammography plus tomosynthesis were 2.24 times those for mammography alone (3.53 mGy and 1.58 mGy, respectively). However, several approaches for reducing dose are being investigated, one of which is the use of synthesized mammographic images reconstructed from the three-dimensional data set (38).

Our study has several limitations. First, there are four interpretation arms in the trial; however, our results are not affected by the other experimental arms as these were based on a comparison of relative performance levels. We included additional arms for investigating the use of computer-aided detection (not used clinically in our practice) and the possible use of synthesized images in lieu of original dose requiring images. The rudimentary reconstructed or synthesized mammographic images we used had been developed 3 years previously and are not current state of the art, and they were not assessed in this study. In addition, tomosynthesis was not assessed as a stand-alone modality in our study. Second, our study included 50–69-year-old women screened biennially, resulting in a higher cancer detection rate than that expected during annual screening. This practice is in accordance with the European Guidelines but is different from that implemented in the United States. However, the average cancer size, histologic findings, and grade distributions at detection suggest that our inferences may be applicable to other practices. Third, our consensus-based arbitration step could have decreased preferentially the recall rates of women suspected of having an abnormality (rating ≥ 2) with only one modality whose case was later dismissed during arbitration; however, to date, we have found no evidence that interval cancers would have changed any of our conclusions. Fourth, during specific days, all women scheduled for screening were approached for possible participation; however, we cannot exclude a possible self-selection bias because we were not allowed to record the reasons why some women declined to participate. Fifth, despite our efforts, we could not completely balance the interpretation load for each of the readers in all four modalities in a busy clinical environment because some radiologists are not at the clinic full time. This situation required us to adjust our analyses for individualized performance levels. Last, this is a single-institution study with a single group of radiologists. However, considering the size of the study and the interreader performance variability combined with the magnitude of the observed performance differences, we believe that similar gains are likely to occur across different practices.

In conclusion, we found a significant increase in cancer detection rates, particularly for invasive cancers, and a simultaneous decrease in false-positive rates with use of mammography plus tomosynthesis compared with mammography alone.

Acknowledgment: Tomosynthesis equipment and support for additional readings were provided by Hologic.

Disclosures of Conflicts of Interest: P.S. Financial activities related to the present article: received financial support for additional overtime readings from Hologic; institution received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **A.I.B.** Financial activities related to the present article: institution received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **R.G.** No relevant conflicts of interest to disclose. **E.B.E.** Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose.

U.E. Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **U.H.** Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **M.I.** Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **I.N.J.** Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **G.J.** Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **M.K.** Financial activities related to the present article: received financial support for additional overtime readings from Hologic. Financial activities not related to the present article: none to disclose. Other relationships: none to disclose. **L.T.N.** Financial activities related to the present article: is employed by Hologic. Financial activities not related to the present article: none to disclose. Other relationships: is employed by Hologic. **S.H.** No relevant conflicts of interest to disclose. **D.G.** Financial activities related to the present article: none to disclose. Financial activities not related to the present article: none to disclose. Other relationships: is an employee of Hologic.

References

1. Paap E, Holland R, den Heeten GJ, et al. A remarkable reduction of breast cancer deaths in screened versus unscreened women: a case-referent study. Cancer Causes Control 2010;21(10):1569–1573.
2. Tabár L, Vitak B, Chen THH, et al. Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011;260(3):658–663.
3. US Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;151(10):716–726, W-236.
4. Hendrick RE, Helvie MA. Mammography screening: a new estimate of number needed to screen to prevent one breast cancer death. AJR Am J Roentgenol 2012;198(3):723–728.
5. Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA 2008;299(18):2151–2163.
6. Houssami N, Ciatto S. The evolving role of new imaging methods in breast screening. Prev Med 2011;53(3):123–126.
7. Rhodes DJ, Hruska CB, Phillips SW, Whaley DH, O'Connor MK. Dedicated dual-head gamma imaging for breast cancer screening in women with mammographically dense breasts. Radiology 2011;258(1):106–118.
8. Sickles EA. The use of breast imaging to screen women at high risk for cancer. Radiol Clin North Am 2010;48(5):859–878.
9. Baker JA, Lo JY. Breast tomosynthesis: state-of-the-art and review of the literature. Acad Radiol 2011;18(10):1298–1310.
10. Diekmann F, Bick U. Breast tomosynthesis. Semin Ultrasound CT MR 2011;32(4):281–287.
11. Helvie MA. Digital mammography imaging: breast tomosynthesis and advanced applications. Radiol Clin North Am 2010;48(5):917–929.
12. Rafferty EA. Breast tomosynthesis. Semin Breast Dis 2006;9(3):111–118.
13. Niklason LT, Christian BT, Niklason LE, et al. Digital tomosynthesis in breast imaging. Radiology 1997;205(2):399–406.
14. U.S. Food and Drug Administration. Medical Devices. Selenia Dimensions 3D System - P080003. http://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/DeviceApprovalsandClearances/Recently-ApprovedDevices/ucm246400.htm. Published March 9, 2011.
15. Bernardi D, Ciatto S, Pellegrini M, et al. Prospective study of breast tomosynthesis as a triage to assessment in screening. Breast Cancer Res Treat 2012;133(1):267–271.
16. Gennaro G, Toledano A, di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. Eur Radiol 2010;20(7):1545–1553.
17. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. AJR Am J Roentgenol 2009;193(2):586–591.
18. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. AJR Am J Roentgenol 2007;189(3):616–623.
19. Svahn T, Andersson I, Chakraborty D, et al. The diagnostic accuracy of dual-view digital mammography, single-view breast tomosynthesis and a dual-view combination of breast tomosynthesis and digital mammography in a free-response observer performance study. Radiat Prot Dosimetry 2010;139(1-3):113–117.
20. Teertstra HJ, Loo CE, van den Bosch MAAJ, et al. Breast tomosynthesis in clinical practice: initial results. Eur Radiol 2010;20(1):16–24.
21. 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(3):788–796.
22. 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. Eur Radiol 2008;18(12):2817–2825.
23. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading—Oslo I Study. Radiology 2003;229(3):877–884.
24. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program—the Oslo II Study. Radiology 2004;232(1):197–204.
25. Skaane P, Gullien R, Bjørndal H, et al. Digital breast tomosynthesis (DBT): initial experience in a clinical setting. Acta Radiol 2012;53(5):524–529.
26. Starr SJ, Metz CE, Lusted LB, Goodenough DJ. Visual detection and localization of radiographic images. Radiology 1975;116(3):533–538.
27. Dachman AH, Obuchowski NA, Hoffmeister JW, et al. Effect of computer-aided detection for CT colonography in a multireader, multicase trial. Radiology 2010;256(3):827–835.
28. Guittet L, Bouvier V, Mariotte N, et al. Comparison of a guaiac and an immunochemical faecal occult blood test for the detection of colonic lesions according to lesion type and location. Br J Cancer 2009;100(8):1230–1235.
29. Gutman S, Kessler LG. The US Food and Drug Administration perspective on cancer biomarker development. Nat Rev Cancer 2006;6(7):565–571.
30. Schatzkin A, Connor RJ, Taylor PR, Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees: example of automated cytometry for early detection of cervical cancer. Am J Epidemiol 1987;125(4):672–678.
31. Pepe MS, Alonzo TA. Comparing disease screening tests when true disease status is ascertained only for screen positives. Biostatistics 2001;2(3):249–260.
32. Perry N, Broeders M, de Wolf C, eds. European guidelines for quality assurance in breast cancer screening and diagnosis. 4th ed. Luxembourg: Office for Official Publications of the European Communities, 2006.
33. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977;64(2):191–199.
34. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann Intern Med 2012;156(7):491–499.
35. Kopans DB, Smith RA, Duffy SW. Mammographic screening and "overdiagnosis". Radiology 2011;260(3):616–620.
36. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev 2011;33(1):111–121.
37. Zuley ML, Bandos AI, Abrams GS, et al. Time to diagnosis and performance levels during repeat interpretations of digital breast tomosynthesis: preliminary observations. Acad Radiol 2010;17(4):450–455.
38. 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;19(2):166–171.