Combined Screening with Ultrasound and Mammography Compared to Mammography Alone in Women at Elevated Risk of Breast Cancer: Results of the First-Year Screen in ACRIN 6666

Wendie A. Berg, MDPhD[1,†], **Jeffrey D. Blume, PhD**[2], **Jean B. Cormack, PhD**[2], **Ellen B. Mendelson, MD**[3], **Daniel Lehrer, MD**[4], **Marcela Böhm-Vélez, MD**[5], **Etta D. Pisano, MD**[6], **Roberta A. Jong, MD**[7], **W. Phil Evans, MD**[8], **Marilyn J. Morton, DO**[9], **Mary C. Mahoney, MD**[10], **Linda Hovanessian Larsen, MD**[11], **Richard G. Barr, MD, PhD**[12], **Dione M. Farria, MD, MPH**[13], **Helga S. Marques, MS**[2], **Karan Boparai, RT(R)(M)**[14], and **the ACRIN 6666 Investigators**[15]

[1] American Radiology Services, Inc., Johns Hopkins Green Spring, Lutherville, MD

[2] Center for Statistical Sciences, Brown University, Providence, RI

[3] Feinberg School of Medicine, Northwestern University, Chicago, IL

[4] CERIM, Buenos Aires, Argentina

[5] Weinberg Imaging Associates, Pittsburgh, PA

[6] University of North Carolina School of Medicine, Chapel Hill, NC

[7] University of Toronto, Sunnybrook and Women's Hospital, Toronto, Ontario, Canada

[8] University of Texas Southwestern Medical Center, Dallas, TX

[9] Mayo Clinic, Rochester, MN

[10] University of Cincinnati, Cincinnati, OH

[11] Keck School of Medicine, University of Southern California, Los Angeles, CA

[12] Forum Health, Western Reserve Care System, Youngstown, OH

[13] Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, MO

[14] American College of Radiology, Philadelphia, PA

Abstract

Context—Screening ultrasound (US) may depict small, node-negative breast cancers not seen on mammography (M).

†Corresponding author WAB: 10755 Falls Road, Suite 440, Lutherville, MD 21093, phone 410-583-2705 or 410-917-3345 (cell), fax 410-616-2812, e-mail: E-mail: wendieberg@gmail.com.
[15]Detailed in Appendix.

Objective—To compare the diagnostic yield (proportion of women with a positive screen test and positive reference standard) and performance of screening with US+M compared to M alone in women at elevated risk of breast cancer.

Design, Setting, and Participants—From April 2004 to February 2006, 2809 women at elevated risk for breast cancer, with at least heterogeneously dense breast tissue in at least one quadrant, were recruited from 21 IRB-approved sites to undergo mammography (M) and physician-performed ultrasound (US) exams in randomized order by a radiologist masked to the results of the other exam. Reference standard was defined as a combination of pathology and 12 month follow-up, and was available for 2637 out of the 2725 eligible participants.

Main Outcome Measure—Diagnostic yield, sensitivity, specificity, and AUC of combined M +US compared to M alone; PPV of biopsy recommendations for M+US compared to M alone.

Results—Forty participants (41 breasts) were diagnosed with cancer: 8 suspicious on both US and M, 12 on US alone, 12 on M alone, and 8 participants (9 breasts) on neither (interval cancers). The diagnostic yield for M was 7.6 per 1000 women screened (20/2637) and increased to 11.8 per 1000 (31/2637) for combined US+M; the supplemental yield was 4.2 per 1000 women screened (95% CI 1.1 to 7.2 per 1000; p = 0.003 that the supplemental yield is zero). The diagnostic accuracy (AUC) for M was 0.78 (95% CI 0.67 to 0.87) and increased to 0.91 (95% CI 0.84 to 0.96) for US+M (p = 0.003 that difference is zero). Of 12 supplemental cancers seen only by US, 11 (92%) were invasive with median size 10 mm (range 5 to 40 mm; mean 12.6, SE 3.0) and 8/9 (89%) reported had negative nodes. PPV of biopsy recommendation after full diagnostic workup (PPV2) was 84/276 for M (22.6%, 95% CI 14.2 to 33%), 21/235 for US (8.9%, 95% CI 5.6 to 13.3%), and 31/276 for combined US+M (11.2%, 95% CI 7.8 to 15.6%).

Conclusions—Adding a single screening US to M will yield an additional 1.1 to 7.2 cancers per 1000 high-risk women, but will also substantially increase the number of false positives. Evaluation of the role of annual screening US is ongoing in this patient population. [Clinicaltrials.gov registry # NCT00072501]

Introduction

Early detection has been proven to reduce deaths due to breast cancer. The United States Preventive Services Task Force analyzed results from 7 randomized trials of mammographic screening and the point estimate of the reduction in mortality from screening mammography was 22% (95% confidence interval [CI] 13 to 30%) in women 50 years of age or older and 15% (95% CI 1 to 27%) among women 40–49 years of age,[1] with some individual trials showing far greater benefits in both age groups. The magnitude of reduction in mortality seen in individual trials parallels reductions in size distribution[2] and rates of node-positive breast cancer.[3]

Mammography can depict calcifications due to malignancy, including ductal carcinoma in situ (DCIS). Invasive cancers, which can spread to lymph nodes and cause systemic metastases, are most often manifest as noncalcified masses,[4] and can be mammographically subtle or occult, particularly when the parenchyma is dense. Dense breast tissue is common, with over half of women under age 50[5] having either "heterogeneously dense" (visually estimated as 51–75% glandular[6]) or "extremely dense" (visually estimated as > 75% glandular[6]) breasts, as do at least one third of women over age 50.[5] In women with dense breasts, mammographic sensitivity may be as low as 30–48%,[7, 8] with much higher interval cancer rates[7, 9] and worse prognosis for resulting clinically detected cancers. Further, dense breast tissue is itself a marker of increased risk of breast cancer on the order of 4- to 6-fold.[10] In dense breasts, digital mammography has improved performance, with sensitivity increasing from 55% with screen film to 70% with digital in one large series using mammographic and clinical follow-up as a gold standard.[11] Digital mammography does not, however, eliminate the fundamental

limitation that noncalcified breast cancers are often obscured by surrounding and overlying dense parenchyma.

In women < age 50, the reduced benefit of mammographic screening is attributed to increased breast density, biologically more aggressive cancers, and reduced prevalence of disease. Using a screening interval of 12 months, rather than 24 months, should improve results with rapidly-growing malignancies,[12] though dense tissue remains a major limitation to improving outcomes.[12] Methods to address improving detection despite dense breast tissue are needed.

Supplemental screening ultrasound (US) has the potential to depict early, node-negative breast cancers not seen on mammography (M),[8, 13–17] and its performance is improved, if anything, in dense parenchyma.[8] Methods that improve detection of small, node-negative cancers should further reduce mortality when performed in addition to screening mammography (M). Ideally, a randomized controlled trial with mortality as an endpoint would be performed to assess any new breast cancer screening technology. However, such trials are extremely costly, participants will often crossover to the additional testing if they perceive a possible benefit, and the technology has typically changed dramatically by the time any results are available. Surrogate endpoints, such as the size and stage of breast cancers depicted, have been correlated with mortality outcomes,[18, 19] and are independent of method of detection.

Across 42,838 examinations from the six published single-center studies of screening US to date,[8, 13–17] 126 women (0.29%) were shown to have 150 cancers identified only on supplemental US.(summarized in [20]) Of 141 (94%) invasive cancers seen only on US, 99 (70%) were 1 cm or smaller in size.(summarized in [20]) Where staging was detailed, 36/40 (90%) cancers seen only sonographically were stage 0 or I.(summarized in [20])

Concerns remain, however, over the generalizability of such favorable results with screening US. In particular, there is concern for the operator dependence of freehand screening breast ultrasound, as an abnormality must be perceived while scanning for it to be documented. Importantly, recent reports have shown that consistent breast US exam performance and interpretation is possible with minimal training.[21, 22] Other limitations to implementing widespread screening US include a shortage of qualified personnel to perform and interpret the examination and lack of standardized scanning protocols. These concerns have hampered use of screening US; 35% of surveyed facilities specializing in breast imaging offered it in 2005,[23] though most facilities offering it will do so only on a limited basis.

Here we report a prospective, multicenter trial, randomized to sequence of performance of mammography and US, designed to investigate and validate the performance of screening US in conjunction with mammography, using a standardized protocol and interpretive criteria. This trial was designed to compare the diagnostic yield of screening breast US+M to M alone in women at increased risk of breast cancer. Since beginning this trial, one other multicenter study of breast US was published from Italy[24] in which 6449 women with dense breasts and negative mammograms underwent screening US, with 29 cancers detected by US (cancer detection rate 0.45%). Ours is the largest trial of screening US in which mammography and ultrasound have been performed and read independently, allowing detailed analysis of the performance of each modality separately and in combination, and reducing potential biases in patient recruitment and interpretation of both mammography and US. Further, we utilized standardized scanning and interpretive criteria (www.acrin.org/TabID/153/Default.aspx), which should facilitate generalizability of our results.

Unlike previous reports evaluating screening US, we chose to study a population at elevated risk of breast cancer. Supplemental screening in addition to mammography may be more cost effective in such populations due to higher expected prevalence of disease. Further, patients at higher risk may be encouraged to begin screening at an earlier age when the tissue is denser

and mammography is more limited in its benefits. Indeed, annual MRI is now recommended in addition to mammography for women at very high risk of breast cancer,[25] but it remains limited by high cost, required injection of contrast, reduced patient tolerance, and limited availability and expertise. Ultrasound is relatively inexpensive, requires no contrast, is well tolerated and is widely available.

Methods

Study Design

Participants were women at high risk of breast cancer (Table 1) who presented for routine annual mammography and provided written informed consent. Each participant underwent M and US screening exams in randomized order (every other two sequential case numbers were assigned to either US first or M first) with the interpreting radiologist for each exam masked to results of the other. If the recommendation from the study M or US was for other than routine annual screening, an integrated US+M interpretation was recorded by a qualified site investigator radiologist. Otherwise, if both US and M were interpreted as negative or benign, no separate integrated interpretation was performed, and the combination of US+M was assumed to be negative. Management was based on recommendations from the integrated exam. If needed, targeted US or additional mammographic views were then performed, and results, assessments, and recommendations, separately recorded. Results of repeat screening at 12 and 24 months after study entry are still being collected.

Web-based data capture and quality monitoring was conducted by ACRIN's Biostatistics and Data Management Center (BDMC). Data were cleaned and locked as of May 14, 2007 for all analyses in this manuscript. The study received Institutional Review Board approval at all participating sites, National Cancer Institute-Cancer Imaging Program approval, and Data and Safety Monitoring Committee review every six months.

Participant Population

2809 women were recruited from 21 sites between April 2004 and February 2006, of whom 2725 were eligible (Fig. 1, Table 1). Women at least 25 years of age who presented for routine annual mammography were eligible to participate if they met uniform definitions of elevated risk (Table 1) as determined by study personnel, and had heterogeneously dense or extremely dense parenchyma[6] in at least one quadrant, either by prior mammography report or review of prior mammograms. Otherwise eligible women with no prior mammography were allowed to enroll since it was felt that such women would be high-risk young women presenting for baseline screening, who would usually have dense breasts. Women were excluded if they had signs or symptoms of breast cancer, recent surgical or percutaneous image-guided breast interventional procedures or magnetic resonance imaging (MRI) or tomosynthesis of the breast (s) within the prior 12 months, or mammography or whole breast US fewer than 11 months earlier. Also excluded were women with breast implants and those who were pregnant, lactating, or planning to become pregnant within two years of study entry or had known metastatic disease. We did not exclude women with prior breast cancer or basal or squamous cell skin cancer or in situ cervical cancer. Women with other prior cancers were eligible to enroll if they had been disease-free for $\geq$ 5 years.

Screening Methods

At least two-view mammography was performed using either screen-film or digital mammography. Visually estimated overall mammographic breast density on study mammograms was recorded as < 25%; 26–40%; 41–60%; 61–80%; or > 80% dense. Computer-assisted detection was not permitted. Radiologist investigators who had successfully completed both phantom scanning[26] and mammographic and sonographic interpretive skills tasks[27]

performed separate, masked interpretations of mammography and US examinations. Survey US was performed using high-resolution linear array, broad bandwidth transducers with maximum frequency of at least 12 MHz, with scanning in transverse and sagittal planes. Lesions other than simple cysts were imaged with and without spatial compounding and power or color Doppler in orthogonal planes (typically radial and antiradial orientations). An image (with embedded clock time) was recorded on entering the ultrasound suite, at the beginning and end of sonographic screening, and on leaving the suite, to determine the time to scan and the total physician time in the room. Electively, the axilla could be scanned, and its inclusion was recorded. Investigators recorded sonographic background echotexture and lesion features using BI-RADS:US descriptors[28] and average breast thickness to the nearest cm.

Assessments for each lesion, and for each breast overall, were recorded on the expanded 7-point BI-RADS[6] scale: 1, negative; 2, benign; 3, probably benign; 4a, low suspicion; 4b, intermediate suspicion; 4c, moderate suspicion; and 5, highly suggestive of malignancy. To facilitate ROC analysis, we did not allow use of BI-RADS[6] 0, though we did allow a recommendation for additional imaging. Investigators were also asked to rate likelihood of malignancy (from 0 to 100%). Recommendations for routine annual follow-up, short interval follow-up in 6 months, additional imaging, and biopsy, were recorded separately from assessments.

Determination of Reference Standard

Reference standard information is a combination of biopsy results within 365 days and clinical follow-up at one year. One year follow-up was targeted for 365 days after the last screening date and very few visits were early; 1.2% (32) occurred before 11 months and 0.46% (12) before 10.5 months. The absence of a known diagnosis of cancer on a participant interview and/or review of medical records at one-year follow-up screen was considered "disease negative" as were two cases with double prophylactic mastectomies. Biopsy results showing cancer (in situ or infiltrating ductal carcinoma, or infiltrating lobular carcinoma) in the breast or axillary lymph nodes were considered malignant, "disease positive," as was one "other invasive" cancer which proved to be a case of melanoma metastatic to axillary lymph nodes. The melanoma case and other non-breast cancers have been systematically excluded from future analyses, but it is retained herein to avoid biasing results as initially reported by the sites. Excision was prompted for core biopsy results of atypical or high-risk lesions including atypical ductal or lobular hyperplasia, lobular carcinoma in situ (LCIS), atypical papilloma, and radial sclerosing lesion.

Statistical considerations

Statistical software (SAS, version 9.1, SAS Institute, Cary, NC; STATA, version 8, Stata Corporation, College Station, TX; S-PLUS, version 7, Insightful Corporation, Seattle, WA and ROCKIT, version 0.9.4 beta [available from the Kurt Rossmann Laboratories for Radiologic Image Research at the University of Chicago, Chicago, Ill, at http://wwwradiology.uchicago.edu/krl/index.htm]) was used in the statistical analysis.

The primary unit of analysis is the participant, with the most severe breast imaging assessment used as the endpoint. A BI-RADS assessment of 4a, 4b, 4c, or 5 was considered "positive" ("seen", "suspicious") for the M or US imaging test or combination of tests, and an assessment of BI-RADS 1, 2, or 3 was considered "negative", as is standard in audits of mammography outcomes.[6, 29] We separately analyzed results based on recommendations, with additional imaging or biopsy or both considered "positive" and short interval or routine follow-up considered "negative". Sample size projections were designed to achieve both the desired level of statistical precision for estimating the yields and at least 80% power to detect a difference in the yields of at least 3 per 1,000, while allowing for 17% missing data.

The diagnostic yield (proportion of women with a positive screen test and positive reference standard), sensitivity, specificity, PPV and NPV were estimated as simple proportions with exact 95% confidence intervals. McNemar's test was used to compare proportions (because of pairing within a participant) and inverted to provide a CI for their difference. Conditional logistic regression was also used. Comparison of PPVs and NPVs was done according to Leisenring et al[30]. For sensitivity at the lesion level, we accounted for clustering by using a logistic regression with robust standard errors. Empirical and model-based ROC curves were estimated from degree of suspicion (BI-RADS) and quasi-continuous probability scales pooled across the study.[31] The areas under the curve (AUCs) were compared under a bivariate, binormal model that accounts for the paired test design.[32, 33]

Of 2725 eligible participants enrolled, only 3.23% (88) were excluded due to missing data. Thirteen (0.48%) never completed imaging and 75 (2.75%) yielded no reference standard information (Fig. 1). The analysis cohort, consisting of all eligible participants with assessment data and reference standard (n=2637), was compared to the full eligible study cohort (n=2725) on baseline characteristics to detect potential biases (Table 1). We note that among the 88 participants with missing data, we would expect only one cancer if the data are missing at random.

Results

The group profile of the 2637 participants (Fig. 1) in our analysis cohort (4786 breasts) was representative of the eligible group (Table 1). Mean age at enrollment was 55 years (SE 0.2, range 25–91). Fourteen hundred women (53%) had a personal history of breast cancer, as did another nine of 23 BRCA-1 or -2 mutation carriers and four of eight women with prior chest/ mediastinal radiation. Seventy-three percent of participants had had prior mammography from ≥11 full months to≤14 months prior to study entry; 11% had prior screening US and 7% had prior contrast-enhanced breast magnetic resonance imaging (MRI) at least one year prior to study entry.

Forty of 2637 participants (1.5%) were diagnosed with cancer, including 39 with breast cancer: DCIS in six, invasive ductal carcinoma (IDC±DCIS) in 20, invasive lobular carcinoma in three, and mixed invasive ductal and lobular carcinoma (±DCIS) in ten. One participant had melanoma metastatic to axillary nodes with no evidence of cancer in the breasts. One patient with IDC had contralateral DCIS (41 total breasts with cancer). Four patients had multifocal invasive cancer (45 total malignant lesions). Median size of invasive cancers (considering only the largest per participant) was 12.0 mm (range 4 to 40 mm, IQR 8 to 18 mm, mean 14, SE 1.5, 95% CI, 11 to 17). Axillary lymph node staging was performed for 25 participants with invasive cancer, with nodal metastases found in five (20%, including the melanoma); axillary staging was not performed for those patients with recurrent breast cancer.

At the participant level, based on BI-RADS assessments, 50% of cancers (20/40) were identified on M (Tables 2 and 3), for a yield of 7.6 per 1000; 5/6 (83%) DCIS were seen only on M. Fifteen invasive cancers, with median size 12 mm (range 4 to 25 mm, IQR 7 to 20 mm, mean 14, SE 1.9, 95% CI, 9.9 to 18.2) were seen on M, with axillary nodes negative in 7/10 (70%) with staging. Seven invasive cancers were suspicious only on M and eight were suspicious on both M+US. Twelve participants had cancer seen only by US: one DCIS and 11 invasive cancers with median size 10 mm (range 5 to 40 mm, IQR 6 to 15 mm, mean 12.6, SE 3.0, 95% CI, 6 to 19), with axillary nodes negative in 8/9 (89%) with staging. One 4-mm IDC considered suspicious initially on M (true positive on M) was downgraded to BI-RADS 3 after integration with US (false negative on M+US), though it was still recalled for additional mammographic views (felt to be probably benign after recall, and benign at 6 month follow-

up), and was diagnosed when the patient presented with palpable metastatic adenopathy 264 days after study entry; this is not included among interval cancers.

The yield of combined US+M was 11.8 per 1000 (31/2637), for a supplemental yield due to US of 4.2 per 1000 (95% CI, 1.1 to 7.2; Table 2). The diagnostic accuracy (AUC) of M alone was 0.78 (95% CI, 0.67 to 0.87), for US alone was 0.80 (95% CI, 0.70 to 0.88), and for combined US+M was 0.91 (95% CI, 0.84 to 0.96, Table 2, Fig. 2). The AUC for US+M did not change when incorporating full diagnostic workup that included additional mammographic views.

Defined as the percent of participants with a BI-RADS 4a assessment or higher, without a diagnosis of cancer in the following 12 months, the false positive rate for M alone was 4.4% (116/2637, Table 3, 95% CI 3.7 to 5.3), for US alone was 8.1% (213/2637, 95% CI 7.1 to 9.2) and for combined M + US was 10.4% (275/2637, 95% CI 9.3 to 11.7). In 5.2% of participants (136/2637; 95% CI, 4.3 to 6.1%), US, but not mammography, resulted in a suspicious assessment and biopsy, and 8.8% (12/136; 95% CI 4.6 to 14.9%) of these participants had cancer.

Table 4 details the recommendations by modality. The PPV1[6] of recall (participants with cancer divided by those recalled for additional evaluation or biopsy or both) was 21/276 (7.6%, 95% CI 4.8 to 11.4%) for mammography, compared to 22/337 (6.5%, 95% CI 4.1 to 9.7%) for US, and 32/436 (7.3%, 95% CI 5.1 to 10.2%) after combined integrated M+US interpretation. Of those 276 participants recalled from routine mammography, after complete diagnostic workup, 84 participants were recommended for biopsy, of whom 19 (PPV2[6], 22.6%, 95% CI 14.2 to 33%) proved to have cancer. PPV2 for US biopsy recommendation after workup was 21/235 (8.9%, 95% CI 5.6 to 13.3%), with one of these cancers classified as BI-RADS 3 on initial US but still worked up, and BI-RADS 4b on mammography. PPV2 after full diagnostic workup and both M+US was 31/276 (11.2%, 95% CI 7.8 to 15.6%).

Short Interval Follow-Up and BI-RADS 3

Of 2637 participants, 177 (6.7%) were classified as BI-RADS 3 on mammography (Table 3); of the 177, one (0.6%) was diagnosed with cancer seen only on US (at early second screen) 363 days after study entry (after initial additional mammographic recall at time zero for unrelated benign findings). Of 2637 participants, 321 (12.2%) were classified as BI-RADS 3 on screening US. Of the 321 considered BI-RADS 3 on initial US, five (1.6%) were diagnosed with cancer within the first 12 months of follow-up: three had suspicious findings on initial mammography and two interval cancers were identified incidentally as a result of six-month follow-up ultrasound for complicated cysts (one a 7 mm IDC found at surgery in adjacent tissue after a core biopsy result of LCIS from the lesion being followed and the other a 27 mm IDC-DCIS adjacent to the cyst being followed; both participants were node negative). Short interval follow-up was recommended for 59/2637 (2.2%, 95% CI 1.7 to 2.9%) participants based on M, for 227 (8.6%, 95% CI 7.6 to 9.7%) participants based on US (with 220 of these based only on US), and for 286 (10.8%, 95% CI 9.7 to 12.1%) of participants after combined M+US.

Twenty-seven participants initially assessed as BI-RADS 3, seven as BI-RADS 4a, and one as BI-RADS 4b on mammography were downgraded to BI-RADS 2 after integration with US findings. There were 26 participants initially considered BI-RADS 3, three as BI-RADS 4a, four as BI-RADS 4b and one as BI-RADS 5 on US which were downgraded to BI-RADS 2 after integration with mammography.

Interval Cancers

Eight participants had cancer not considered suspicious on either M or US, with cancer identified during the 12 months after initial screening, i.e. "interval cancers". Three node-

negative cancers (35 mm ILC; 8 mm IDC+ILC; and a 20 mm IDC-DCIS) were identified at the second screen (performed early, after 11 full months), with biopsies at from 359 to 364 days after study entry. One participant noted a palpable lump, with biopsy showing 12 mm mixed IDC/ILC 337 days after study entry. One participant presented with skin recurrence of prior breast cancer 231 days after study entry. Two cancers were found at six-month follow-up ultrasound as detailed in the section on short interval follow-up. One non-breast malignancy was identified in the interval in a participant with prior melanoma of the back, who, 6 years later, developed palpable axillary mass due to metastatic adenopathy, with no evidence of malignancy within the breasts. Thus, the interval cancer rate was 8/40 (20%) if the melanoma case is included as cancer, or 7/39 (18%) if not; only 2/39 (5.1%) of participants with breast cancer were identified due to symptoms in the interval between screenings [or 3/39 (7.8%), if one includes the 4-mm IDC seen on initial M but not on additional imaging or at 6-month follow-up, which presented with palpable metastatic adenopathy]. A ninth breast had cancer not seen on either M or US: DCIS was identified only at prophylactic mastectomy after diagnosis of contralateral multifocal IDC seen only on US.

Cancers seen only on US were evenly distributed across breast density categories (Table 5). The data were inconclusive with respect to most differences between film-screen and digital mammography (Table 5); however, slightly higher specificity was observed with digital mammography than film screen (97.0% vs. 94.7%, p=0.007).

In 1400 women with a personal history of breast cancer, 28 (2.0%) were found to have cancer, with 9/28 (32%) seen only on US (Table 5); cancers were evenly distributed between the breast ipsilateral to the initial cancer and contralateral disease. Among 1237 women with risk factors other than a personal history of breast cancer, 12 (1.0%) were found to have cancer, with 3/12 (25%) cancers seen only on US. Significantly more cancers overall were found in women with a personal history of cancer (p=.03), but there was no difference in supplemental yield of US in women with or without a personal history of breast cancer.

The median time to perform screening breast US was 19 minutes (range 2 to 90 min; IQR 12 to 27, mean 20.8, SE 0.3, 95% CI 20.3 to 21.3) for a bilateral scan and 9 minutes for a unilateral scan (range 1 to 70 min, IQR 5 to 15, mean 11.6, SE 0.4, 95% CI 10.7 to 12.4). A median of another 2.0 minutes was spent in the room with the participant (range 0 to 19, IQR 2 to 3, mean 2.7, SE 0.04, 95% CI, 2.6 to 2.7). For 869/2637 (33.0%) of participants, the investigator scanned at least one axilla while performing US scanning of the breast(s). Ninety-four percent of breasts were < 4 cm in thickness.

Discussion

Supplemental physician-performed screening US increases the cancer detection yield by 4.2 cancers per 1000 women at elevated risk of breast cancer, as defined in this protocol (95% CI, 1.1 to 7.2 cancers per 1000) on a single, prevalent screen. This is similar to rates of US-only cancers of 2.7 to 4.6 cancers per 1000 women screened in other series.[8, 13–17, 24] As in prior studies, the vast majority of cancers seen only on US were invasive, as DCIS is difficult to see on US. All but one cancer seen only on US was node negative. Invasive cancers not seen on mammography can be expected to present as interval cancers with worse prognosis: detection of asymptomatic, mammographically-occult, node-negative invasive carcinomas with US should reduce mortality from breast cancer, although mortality was not an endpoint of this study.

Strengths of our study include its matching within a participant, and exams performed by radiologists who were masked to results of the other exam. Randomized order of these exams helped control biases of recruiting women with vague mammographic abnormalities. Further,

these results were consistent and generalizable across 21 international centers. The radiologist investigators in this trial were all specialists in breast imaging who met experience requirements and completed qualification tasks. As such, our results may vary slightly from those observed in general practice, though similar results were observed by Kaplan[16] where technologists performed screening US. Educational materials used for radiologist investigator training in US lesion detection and characterization are archived by ACRIN.

The use of the Gail and Claus models to calculate risk may have affected the racial distribution of participants, as the Gail model is known to underestimate risk in African Americans.[34] Neither model has been validated in other races other than Caucasians,[34, 35] though Gail et al. [36] have recently validated a new risk assessment tool based on data from the Contraceptives and Reproductive Experiences (CARE) Study in African American women (which was not available for use in this protocol).

In our elevated-risk study population, enriched in women with dense breasts, mammographic sensitivity was only 50% (95% CI, 33.8 to 66.2) and the sensitivity of US+M was 77.5% (95% CI, 61.55 to 89.16) (Table 2). From a detection standpoint, it may be reasonable to offer supplemental screening US to women with similar risk criteria. As stated, dense breast tissue is common: approximately half of women under age 50 and a third of older women have dense breast parenchyma.[5] Approximately 6% of women presenting for routine annual mammography have a personal history of breast cancer[29] and 15% have a family history of breast cancer.[29]

Our ongoing study, allowing for contrast-enhanced breast magnetic resonance imaging (MRI) within 8 weeks of the final 24 month mammography/US follow-up round, may shed some light on the possible competitive roles of US and MRI as adjuncts to mammographic screening for breast cancer. Across four other series where mammography, US, and MRI had been performed for screening women at very high risk of breast cancer, combined sensitivity of mammography and US averaged 55%, compared to 93% after combined mammography and MRI.[37–40] There appears to be no role for screening US in women undergoing screening MRI, though US may be helpful in guiding biopsy of suspicious findings seen first on MRI.[37–40] US may be more appropriate than MRI for screening women of intermediate risk due to its reduced cost relative to MRI. Many of the cancers seen only on MRI are small, node-negative invasive cancers. [37–40] Unlike US, MRI readily depicts DCIS,[41] although DCIS remains overrepresented among false-negative MRI examinations.[42] It is uncertain whether detection of DCIS is required, or whether detection of node-negative invasive breast cancer is sufficient for a screening test. It will be important to see the stage distribution of breast cancers in subsequent rounds of screening with US+M in this study, and to know how many invasive cancers will be seen only on MRI at the 24 month time point.

Despite a 20% interval cancer rate (8/40 participants with cancer) in our series, none of the interval breast carcinomas were node-positive; the only interval cancer that was node positive was a non-breast cancer (melanoma metastatic to axillary nodes). Another cancer considered suspicious on initial mammography (and therefore not included among "interval cancers") was considered probably benign after full diagnostic workup and went unbiopsied until the patient presented with palpable, metastatic nodes, yet was only 4 mm in size at eventual detection. One interval cancer was a skin recurrence of prior breast cancer.

US is well tolerated, the technology is widely available, and it does not require intravenous contrast material. If, however, screening US is to be widely implemented, several major issues remain. First, it will be very important to know the role of annual screening US, and such study is in progress with participants in this protocol. The time to perform bilateral screening US is problematic, at a median of 19 minutes. This does not include comparison to prior studies,

discussion of results with patients, nor creation of a final report, although the time may be artificially prolonged by protocol requirements to measure each lesion other than a simple cyst in two planes, and to fully characterize each such lesion with and without spatial compounding and with color or power Doppler. Nineteen minutes is considerably longer than the average 4 minutes 39 seconds reported by Kolb et al[8] for physicians scanning, or the average 10 minutes reported by Kaplan et al[16] for technologists. Currently, there is only a single billing code for breast ultrasound (76645), and Medicare global reimbursement is $85 in 2008, which does not fully cover the costs of performing and interpreting the examination. Results similar to those of our physician-performed study have been reported with technologist-performed US,[16] and specialized training of technologists is encouraged to counter a current shortage of qualified physician and technologist personnel. Further validation of technologist-performed screening breast US is encouraged. Automated whole-breast US may facilitate implementation and profitability of screening US, but will result in hundreds of images to be reviewed and stored, with attendant increased capital and professional costs and potential increased malpractice exposure; validation of such methods is needed. The full costs of screening breast US in this protocol, including the costs of induced additional testing and biopsy, are being analyzed and reported separately.

The final barrier to implementing screening US is the risk of false positives. The performance characteristics of mammography were within accepted ranges (10.5% recalled for additional imaging or biopsy; 3.1% of participants biopsied after full workup, with 23% proving malignant; 2.2% recommended for short interval follow-up). We observed a 5.4% recall rate for US (142/2637 recommended for additional imaging), which may be artificially low in this series as physicians performed the screening US and could directly evaluate lesions in real-time. Of 2637 participants, 233 (8.8%) participants had findings considered suspicious on US [with 136 participants having suspicious findings on US but not mammography and prompting biopsy], and 235 (8.9%) were recommended for biopsy based on US. Only 20/233 (8.6%) of those participants with US-suspicious findings [and 12/136, 8.8% of those with suspicious findings biopsied based only on US], and 21/235 (8.9%) of participants biopsied based on US proved to have cancer. The 8.8% to 8.9% PPV of US-prompted biopsy in our study is similar to the 11% rate seen across prior series.[20, 43] Diagnostic uncertainty for complicated cysts remains a major source of false positives, with 47 participants undergoing only cyst aspiration included among those recommended for biopsy based on US (Table 3). Elastography, in which the deformability of the mass is assessed during US, can help distinguish complicated cysts from suspicious solid masses and should reduce this source of false positives.[44] Another 227 (8.6%) of participants were recommended for short interval follow-up based on US, similar to the 6.3% across other series.[8, 15, 16, 45] It is likely that the risk of false positives with US will diminish with subsequent screening rounds, as has been seen with mammography[46] and, in small series, with both US and MRI[37]; this evaluation is in progress. We have been separately quantifying patient anxiety and discomfort (i.e. "process utility"[47]) induced by addition of screening US.

In summary, the addition of a single screening US examination to mammography in women at elevated risk of breast cancer results in increased detection of breast cancers which are predominantly small and node-negative. We defined "elevated risk" using a variety of criteria, including personal history of breast cancer, prior atypical biopsy, and elevated risk by Gail or Claus models or both. Recent literature[43] suggests that any combination of factors that confers three-fold relative risk compared to women without the risk factor would be "high risk," including dense breast tissue.[9] Across all series to date, over 90% of cancers seen only on US have been in women with > 50% dense breast tissue,[20, 24] though 3/12 (25%) of the cancers seen only on US in this series were in women with only 26–40% dense breast tissue (as visually estimated), suggesting that women with other risk factors may benefit from screening US even if their breast tissue is less dense. The age at which to begin screening women at increased risk

would reasonably derive from the age at which the risk of breast cancer is equal to that for an average woman of age 40 or 50, depending on national policy.[9]

The detection benefit of a single screening US in women at elevated risk of breast cancer is now well validated. However, it comes with a substantial risk of false positives (i.e. biopsy with benign results and/or short interval follow-up). Our results should be interpreted in the context of recent guidelines recommending annual MRI in women at very high risk of breast cancer.[25] Importantly, evaluation of annual (incidence) screening US is continuing in ACRIN 6666, as is evaluation of a single screening MRI in these women.

Acknowledgments

References

1. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med Sep 3;2002 137(5 Part 1):347–360. [PubMed: 12204020]

2. Tabar L, Vitak B, Chen HH, et al. The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. Radiol Clin North Am 2000;38(4):625–651. [PubMed: 10943268]

3. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am Sep;2004 42(5):793–806. v. [PubMed: 15337416]

4. Tabar L, Tot T, Dean PB. Early detection of breast cancer: Large-section and subgross thick-section histologic correlation with mammographic appearances. Radiographics 2007;27:S5–S35.

5. Stomper PC, D'Souza DJ, DiNitto PA, Arredondo MA. Analysis of parenchymal density on mammograms in 1353 women 25–79 years old. AJR Am J Roentgenol Nov;1996 167(5):1261–1265. [PubMed: 8911192]

6. D'Orsi, CJ.; Bassett, LW.; Berg, WA., et al. Breast Imaging Reporting and Data System, BI-RADS: Mammography. Vol. 4. Reston: American College of Radiology; 2003.

7. Mandelson MT, Oestreicher N, Porter PL, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst Jul 5;2000 92(13): 1081–1087. [PubMed: 10880551]

8. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: An analysis of 27,825 patient evaluations. Radiology Oct;2002 225(1):165–175. [PubMed: 12355001]

9. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med Jan 18;2007 356(3):227–236. [PubMed: 17229950]

10. Harvey JA, Bovbjerg VE. Quantitative assessment of mammographic breast density: relationship with breast cancer risk. Radiology Jan;2004 230(1):29–41. [PubMed: 14617762]

11. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med Oct 27;2005 353(17):1773–1783. [PubMed: 16169887]

12. Buist DS, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40–49 years. J Natl Cancer Inst Oct 6;2004 96(19):1432–1440. [PubMed: 15467032]

13. Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dunser M. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. Semin Ultrasound CT MR 2000;21(4):325–336. [PubMed: 11014255]

14. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. AJR Am J Roentgenol Jul;2003 181(1):177–182. [PubMed: 12818853]

15. Gordon PB, Goldenberg SL. Malignant breast masses detected only by ultrasound. A retrospective review [see comments]. Cancer 1995;76(4):626–630. [PubMed: 8625156]

16. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. Radiology Dec;2001 221(3):641–649. [PubMed: 11719658]

17. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. AJR Am J Roentgenol Jun; 2003 180(6):1675–1679. [PubMed: 12760942]

18. Michaelson JS, Silverstein M, Sgroi D, et al. The effect of tumor size and lymph node status on breast carcinoma lethality. Cancer Nov 15;2003 98(10):2133–2143. [PubMed: 14601082]

19. Michaelson JS, Silverstein M, Wyatt J, et al. Predicting the survival of patients with breast carcinoma using tumor size. Cancer Aug 15;2002 95(4):713–723. [PubMed: 12209713]

20. Berg WA. Supplemental screening sonography in dense breasts. Radiol Clin North Am Sep;2004 42 (5):845–851. vi. [PubMed: 15337420]

21. Bosch AM, Kessels AG, Beets GL, et al. Interexamination variation of whole breast ultrasound. Br J Radiol May;2003 76(905):328–331. [PubMed: 12763948]

22. Berg WA, Blume J, Cormack JB, Mendelson EB. Operator dependence of physician-performed whole breast US: Lesion detection and characterization. Radiology 2006;241:355–366. [PubMed: 17057064]

23. Farria DM, Schmidt ME, Monsees BS, et al. Professional and economic factors affecting access to mammography: A crisis today, or tomorrow? Cancer Aug 1;2005 104(3):491–498. [PubMed: 15973693]

24. Corsetti V, Ferrari A, Ghirardi M, et al. Role of ultrasonography in detecting mammographically occult breast carcinoma in women with dense breasts. Radiol Med (Torino) Apr 2006;111(3):440–448.

25. Saslow D, Boetes C, Burke W, et al. American cancer society guidelines for breast screening with MRI as an adjunct to mammography. CA Cancer J Clin Mar-Apr;2007 57(2):75–89. [PubMed: 17392385]

26. Berg WA, Blume JD, Cormack JB, Mendelson EB, Madsen EL. Lesion detection and characterization in a breast US phantom: results of the ACRIN 6666 Investigators. Radiology Jun;2006 239(3):693–702. [PubMed: 16641344]

27. Berg WA, Mendelson EB, Blume JD, Cormack JB. Performance of experienced breast imagers in interpreting breast ultrasound: results of training in the ACRIN 6666 protocol. Amer J Roentgenol 2004;182(4):34 (abstr).

28. Mendelson, EB.; Baum, JK.; Berg, WA.; Merritt, CRB.; Rubin, E. Breast Imaging Reporting and Data System, BI-RADS: Ultrasound. Vol. 1. Reston: American College of Radiology; 2003.

29. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. Radiology Oct;2006 241(1):55–66. [PubMed: 16990671]

30. Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. Biometrics Jun;2000 56(2):345–351. [PubMed: 10877288]

31. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics Sep;1988 44(3):837–845. [PubMed: 3203132]

32. Metz, CE.; Wang, P.; Kronman, HA. New approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck, F., editor. Information processing in medical imaging. The Hague, Netherlands: Nijihoff; 1984.

33. Zhou, X.-H.; Obuchowski, NA.; McClish, DK. Statistical methods in diagnostic medicine. New York: John Wiley; 2002.

34. Adams-Campbell LL, Makambi KH, Palmer JR, Rosenberg L. Diagnostic Accuracy of the Gail Model in the Black Women's Health Study. Breast J Jul-Aug;2007 13(4):332–336. [PubMed: 17593036]

35. Bondy ML, Newman LA. Breast cancer risk assessment models: applicability to African-American women. Cancer Jan 1;2003 97(1 Suppl):230–235. [PubMed: 12491486]

36. Gail MH, Costantino JP, Pee D, et al. Projecting individualized absolute invasive breast cancer risk in African American women. J Natl Cancer Inst Dec 5;2007 99(23):1782–1792. [PubMed: 18042936]

37. Warner E, Plewes DB, Hill KA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. Jama Sep 15;2004 292(11):1317–1325. [PubMed: 15367553]

38. Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. J Clin Oncol Nov 20;2005 23(33):8469–8476. [PubMed: 16293877]

39. Sardanelli F, Podo F, D'Agnolo G, et al. Multicenter comparative multimodality surveillance of women at genetic-familial high risk for breast cancer (HIBCRIT study): interim results. Radiology Mar;2007 242(3):698–715. [PubMed: 17244718]

40. Lehman CD, Isaacs C, Schnall MD, et al. Cancer Yield of Mammography, MR, and US in High-Risk Women: Prospective Multi-Institution Breast Cancer Screening Study. Radiology Aug;2007 244(2):381–388. [PubMed: 17641362]

41. Menell JH, Morris EA, Dershaw DD, Abramson AF, Brogi E, Liberman L. Determination of the presence and extent of pure ductal carcinoma in situ by mammography and magnetic resonance imaging. Breast J Nov-Dec;2005 11(6):382–390. [PubMed: 16297080]

42. Schnall MD, Blume J, Bluemke DA, et al. Diagnostic Architectural and Dynamic Features at Breast MR Imaging: Multicenter Study. Radiology Jan;2006 238(1):42–53. [PubMed: 16373758]

43. Berg WA. Beyond standard mammographic screening: mammography at age extremes, ultrasound, and MR imaging. Radiol Clin North Am Sep;2007 45(5):895–906. vii. [PubMed: 17888776]

44. Booi RC, Carson PL, O'Donnell M, Roubidoux MA, Hall AL, Rubin JM. Characterization of cysts using differential correlation coefficient values from two dimensional breast elastography: preliminary study. Ultrasound Med Biol Jan;2008 34(1):12–21. [PubMed: 17900795]

45. Kolb TM, Lichy J, Newhouse JH. Occult cancer in women with dense breasts: detection with screening US-- diagnostic yield and tumor characteristics. Radiology 1998;207(1):191–199. [PubMed: 9530316]

46. Schell MJ, Yankaskas BC, Ballard-Barbash R, et al. Evidence-based target recall rates for screening mammography. Radiology Jun;2007 243(3):681–689. [PubMed: 17517927]

47. Swan JS, Lawrence WF, Roy J. Process utility in breast biopsy. Med Decis Making Jul-Aug;2006 26(4):347–359. [PubMed: 16855124]

Appendix

Appendix

The following persons served as site investigators (PI) or research associates (RA) at the ACRIN 6666 clinical sites: *Allegheny-Singer Research Institute, Pittsburgh* –William R. Poller, MD (PI), Michelle Huerbin (RA); *American Radiology Services –Johns Hopkins Green Spring, Baltimore, MD* – Wendie A. Berg, MD, PhD (PI), Barbara E. Levit RT(R)(M) (RA); *Beth Israel Deaconess Medical Center, Boston, MA* – Janet K. Baum, MD (PI), Valerie J. Fein-Zachary, MD (PI), Suzette M. Kelleher, BA (RA); *CERIM, Buenos Aires* – Daniel E. Lehrer, MD (PI), Maria S. Ostertag (RA); *Duke University Medical Center, Durham, NC* – Mary Scott Soo, MD (PI), Brenda N. Prince, RT (RA); *Mayo Clinic, Rochester, MN* – Marilyn J. Morton, DO (PI), Lori M. Johnson, AAS (RA); *Feinberg School of Medicine, Northwestern University, Chicago, IL* – Ellen B. Mendelson, MD (PI), Marysia Kalata, AA (RA); *Radiology Associates of Atlanta, Atlanta, GA*– Handel Reynolds, MD (PI), Y. Suzette Wheeler, RN, MSHA, CCRC (RA); *Radiology Consultants/Forum Health, Youngstown, OH* – Richard G. Barr, MD, PhD (PI), Marilyn J. Mangino, RN (RA); *Radiology Imaging Associates, Denver, CO* – A. Thomas Stavros, MD (PI), Margo Valdez (RA); *Sunnybrook Health Sciences Centre, University of Toronto, Toronto* – Roberta A. Jong, MD (PI), Julie H. Lee, BSC (RA); *Thomas Jefferson University Hospital, Philadelphia, PA* – Catherine W. Piccoli, MD (PI), Christopher R. B. Merritt, MS, MD (PI), Colleen Dascenzo (RA); *David Geffen School of Medicine at University of California Los Angeles Medical Center, Los Angeles, CA* –Anne C. Hoyt, MD (PI), Roslynn Marzan, BS (RA); *University of Cincinnati Medical Center, Cincinnati, OH* – Mary C. Mahoney, MD (PI), Monene M. Kamm, AS (RA); *University of North Carolina, Chapel Hill, NC* – Etta D. Pisano, MD (PI), Laura A. Tuttle, MA (RA), *Keck School of Medicine, University of Southern California, Los Angeles, CA* – Linda Hovanessian Larsen, MD (PI), Christina E. Kiss, AA, CCRP (RA); *University of Texas M.D. Anderson Cancer Center, Houston, TX* – Gary J. Whitman, MD (PI), Sharon R. Rice, AA (RA); *University of Texas Southwestern Medical Center, Dallas, TX* – W. Phil Evans, MD (PI), Kimberly T. Taylor, AA (RA); *Washington University School of Medicine, St. Louis, MO* – Dione M. Farria, MD, MPH (PI), Darlene J. Bird, RT(R)(M), AS (RA); *Weinstein Imaging Associates, Pittsburgh, PA* – Marcela Böhm-Vélez (PI), Antoinette Cockroft (RA).

Enrollment and Eligibility for ACRIN 6666

2809
Participants
Enrolled

84 participants ineligible:
- 38 participant medical history not allowable
- 19 risk assessment criteria not met
- 10 duplicate registrations
- 8 enrolled under expired IRB
- 7 diagnostic cases
- 1 metastatic disease at study entry
- 1 breast implants

2725 Eligible
Participants

2712 Participants received complete imaging (US and Mammography)
13 Participants did not receive complete imaging
- 7 Participants withdrew before imaging
- 4 Participants did not receive any imaging
- 2 Participants did not receive US imaging

75 Participants Missing Reference Standard Information:
- 10 alive at 1 year but cancer status unknown
- 8 deceased
- 33 lost to follow-up – no contact at 1 year follow-up
- 24 withdrawn

2637 included in Analysis:
    335 positive
    2302 negative
Matched pair design to compare
    Mammography + US to Mammography alone

Fig. 1.
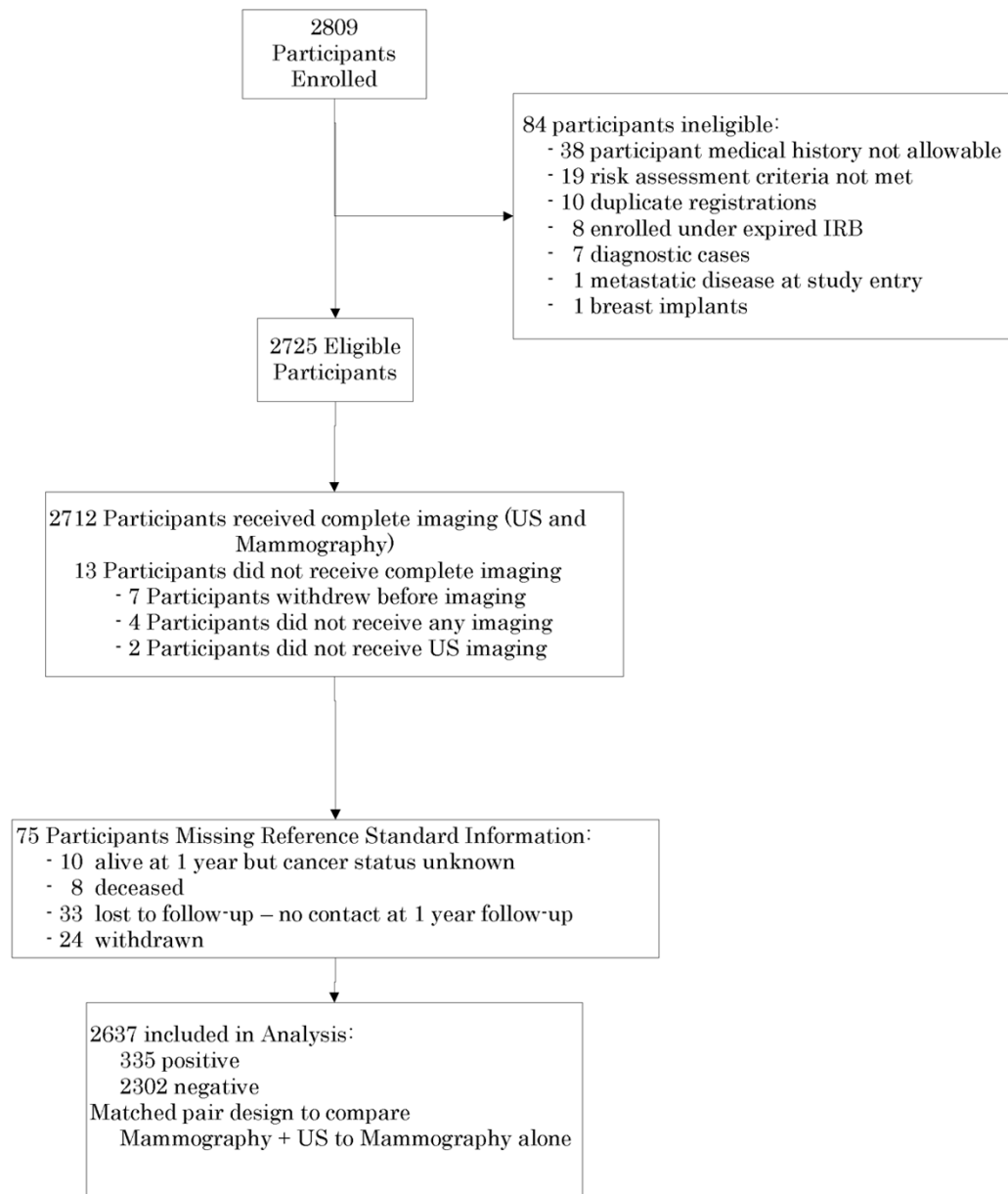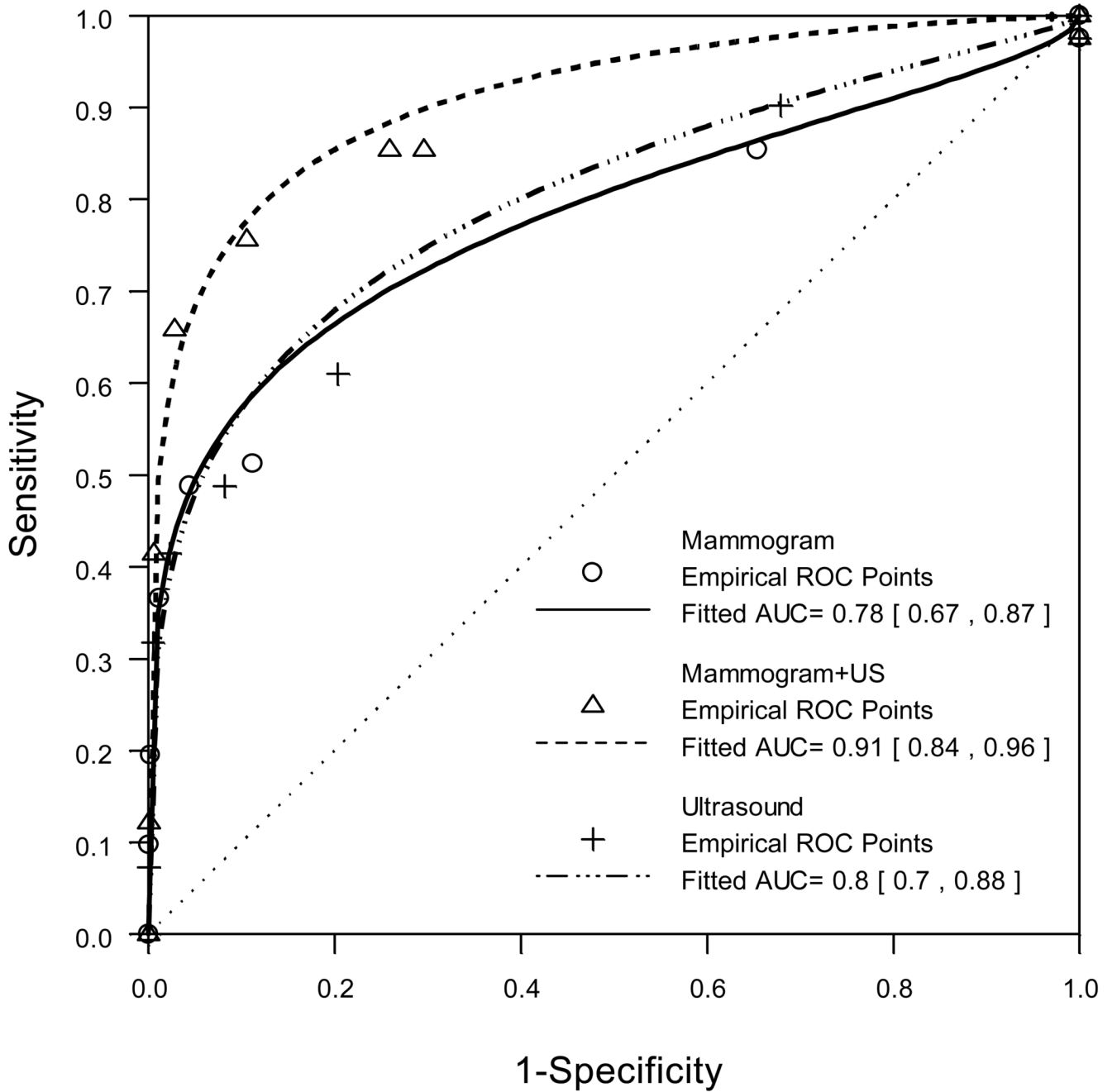Flowchart of protocol.

Fig. 2.
ROC curves based on expanded BI-RADS assessments for mammography alone (solid line), combined mammography and ultrasound (dashed line), and ultrasound alone (dashed and dotted line). Table 2 presents summary statistics for these curves. The ultrasound ROC is included for completeness; the study was not designed to permit direct comparison to ultrasound alone.

Table 1

Participant Characteristics

	Eligible Participants (n = 2,725)	Analysis set (n = 2,637)
Median Age[a] at Enrollment (yrs) (range)	55 (25–91)	55 (25–91)
Mean Age at Enrollment (yrs) (standard error)	55.14 (0.19)	55.23 (0.20)
Race or Ethnicity[b][T]		
White	2519 (92.44)	2448 (92.83)
Hispanic or Latina[c]	274 (10.06)	259 (9.82)
Black or African American	100 (3.67)	90 (3.41)
Native Hawaiian or other Pacific Islander	4 (0.15)	4 (0.15)
Asian	95 (3.49)	88 (3.34)
American Indian or Alaskan Native	4 (0.15)	4 (0.15)
Unknown	11 (0.40)	11 (0.42)
MENOPAUSAL STATUS[d]		
Premenopausal	629 (23.08)	601 (22.79)
Perimenopausal	188 (6.90)	179 (6.79)
Postmenopausal	1387 (50.90)	1353 (51.31)
Surgical Menopause	492 (18.06)	482 (18.28)
Unknown/I cannot remember	22 (0.81)	22 (0.83)
Data missing	7 (0.26)	0 (0.0)
Risk Factor(s)[e]		
Personal History of Breast Cancer	1443 (52.95)	1400 (53.09)
Lifetime Risk ≥ 25% by Gail or Claus Model	517 (18.97)	497 (18.85)
Five Year Risk, Gail Model ≥ 2.5%	411 (15.08)	403 (15.28)
Five Year Risk, Gail Model ≥ 1.7% and Extremely Dense Breasts	230 (8.44)	223 (8.46)
ADH/ALH/LCIS or Atypical Papilloma	84 (3.08)	83 (3.15)
Mutation in BRCA-1 or BRCA-2[f]	24 (0.88)	23 (0.87)
History of Prior Chest and/or Mediastinal and/or Axillary Irradiation[e]	8 (0.29)	8 (0.30)
Basis of eligibility unknown	8 (0.29)	0 (0.0)
Imaging History[T]		
Mammogram < 14 months prior to entry	1977 (72.55)	1929 (73.15)
Mammogram 14–24 months prior to entry	587 (21.54)	561 (21.27)
Mammogram > 24 months prior to entry	107 (3.93)	104 (3.94)
No Prior Mammogram	54 (1.98)	43 (1.63)
Prior breast ultrasound	1891 (69.40)	1834 (69.55)
Prior screening US	304 (11.16)	293 (11.11)
Prior contrast-enhanced breast MRI	196 (7.19)	191 (7.24)
Hormone Use		
Current HRT	148 (5.43)	143 (5.42)
Any prior HRT	792 (29.06)	780 (29.58)
Never HRT	1785 (65.50)	1714 (65.00)

	Eligible Participants (n = 2,725)	Analysis set (n = 2,637)
Current Tamoxifen use at study entry	338 (12.40)	329 (12.48)
Current Raloxifene use at study entry	95 (3.49)	94 (3.56)
Current aromatase inhibitor use at study entry	142 (5.21)	141 (5.35)
BRA CUP SIZE		
A	449 (16.48)	439 (16.65)
B	983 (36.07)	949 (35.99)
C	856 (31.41)	830 (31.48)
D	312 (11.45)	304 (11.53)
DD	91 (3.34)	89 (3.38)
Other, specify	27 (0.99)	26 (0.99)
Data Missing	7 (0.26)	0 (0.0)

Abbreviations used: HRT = hormone replacement therapy.

[a] The age distribution for the eligible set was 43(P10), 48(P25), 55(P50), 62(P75), 68(P90). The age distribution for the analysis set was 43(P10), 49(P25), 55(P50), 62(P75), 68(P90).

[T] Participants may fall into more than one category: percentages (in parentheses) may sum to greater than 100%.

[b] Race and ethnic group were self-assigned.

[c] Of the Hispanic or Latina participants in the eligible and analysis sets, 80% were accrued at CERIM, Buenos Aires, Argentina. None of the Black or African American participants in either set were accrued at CERIM.

[d] Premenopausal women had their last menstrual period less than one month before registration. Perimenopausal women had their last menstrual period at least 1 month but less than 12 months before registration. Postmenopausal women had their last menstrual period more than one year before registration.

[e] Participants are listed only once for a risk factor which determined eligibility. BRCA-mutation carrier status and prior chest/mediastinal radiation were prioritized over personal history of breast cancer, which was prioritized over prior atypical biopsy result, then risk by Gail or Claus models (lifetime risk prioritized over 5 year risks > 2.5% then 1.7% and extremely dense parenchyma respectively).

[f] Nine women with BRCA-1 or -2 mutations also had a personal history of breast cancer, as did four with prior chest and/or mediastinal radiation.

Table 2

Summary of Performance Characteristics of Screening with Combined Mammography and Ultrasound Compared to Mammography Alone at the Participant Level

	Mammography + Ultrasound (M+US)[a]	Mammography (M) Alone	Comparison M+US vs. M	p-value[b]	US Alone[c,d]
Yield (per 1000)	31/2637 11.8 (8, 16.6)	20/2637 7.6 (4.6, 11.7)	Difference 4.2 (1.1, 7.2)	0.003	20/2636 7.6 (4.6, 11.7)
Sensitivity (in %)	31/40 77.5 (61.55, 89.16)	20/40 50 (33.80, 66.2)	Difference 27.5 (9.52, 45.48)	0.003	20/40 50.0 (33.8, 66.2)
Specificity (in %)	2322/2597 89.41 (88.16, 90.57)	2481/2597 95.53 (94.67, 96.30)	Difference −6.12 (−7.24, −5)	<0.001	2383/2596 91.8 (90.7, 92.8)
PPV (in %)	31/306 10.13 (6.99, 14.07)	20/136 14.71 (9.22, 21.79)	Odds Ratio 0.65 (0.36, 1.19)	0.03	20/233 8.6 (5.3, 13.0)
NPV (in %)	2322/2331 99.61 (99.27, 99.82)	2481/2501 99.20 (98.77, 99.51)	Odds Ratio 2.08 (0.95, 4.58)	0.004	2383/2403 99.2 (98.7, 99.5)
Area under ROC curve (BI-RADS)	0.91 (0.84, 0.96)	0.78 (0.67, 0.87)	Difference 0.13 (0.04, 0.22)	0.003	0.80 (0.70, 0.88)
Area under ROC curve (% Prob. Malignancy)	0.90 (0.83, 0.95)	0.68 (0.53, 0.80)	Difference 0.23 (0.10, 0.35)	<0.001	0.75 (0.62, 0.85)

Notes: Estimates are rounded for presentation. Table cells display: raw counts, point estimate and 95% confidence intervals; for the comparison column, the counts are replaced by the metric of comparison. Calculations include one malignancy that was considered suspicious on mammography and was considered probably benign after integration with ultrasound; and one malignancy not seen on initial imaging discovered to be melanoma (from prior primary in back) metastatic to axillary lymph nodes and diagnosed 205 days after study entry.

[a] Table displays screening data. Including complete diagnostic workup gives M+US: Yield = 31/807 (38.4 per 1000, 95% CI 26.2, 54.1); Sensitivity = 31/35 (88.6%, 95% CI 73.3, 96.8); Specificity = 550/772 (71.2%, 95% CI 67.9, 74.4); PPV = 31/253 (12.3%; 95% CI 8.5, 16.9); NPV = 550/554 (99.3%; 95% CI 98.2, 99.8); AUC (BI-RADS) = 0.91 (95% CI 0.84, 0.95); AUC (% Probability of Malignancy) = 0.90 (95% CI 0.83, 0.94).

[b] Testing H₀: M+US = M Alone in stated metric

[c] US results included for completeness; the study was not designed to permit direct comparison to US alone.

[d] For one participant, the US alone interpretation was not available.

Table 3

Sensitivity, Specificity, and PPV by Participant for Mammography Alone (M) vs. Combined Mammography + Ultrasound (M+US)

BI-RADS Assessment	N Participants	N Breast Cancers	Cumulative N Participants	Cumulative No. of True Positives	Sensitivity: All Cancers	Sensitivity: Invasive Cancers	Specificity	PPV: Participant Level	No. of Participants with either a Core or Surgical Biopsy[a]	No. of Lesions with either a Core or Surgical Biopsy	No. of Participants with only a Cyst Aspiration	No. of Lesions with only a Cyst Aspiration	No. of Participants with an Atypical[b] Biopsy	No. of Atypical Lesions that were Biopsied
5	5	4	5	4	0.1000	0.0882	0.9996	0.8000	5	5	0	0	0	0
4C	9	4	14	8	0.2000	0.2059	0.9977	0.5714	9	12	0	1	1	1
4B	32	7	46	15	0.3750	0.3235	0.9881	0.3261	22	24	1	2	0	0
4A	90	5	136	20[c]	0.5000	0.4412	0.9553	0.1471	33	43	3	6	2	2
3	177	1	313	21	0.5250	0.4706	0.8876	0.0671	16	19	12	16	0	0
2[d]	1421	14	1734	35	0.8750	0.8824	0.3458	0.0202	112	131	37	59	7	7
1	903	5	2637	40	1.0000	1.0000	0.0000	0.0152	70	85	18	29	6	8
5	6	5	6	5	0.1250	0.1176	0.9996	0.8333	6	8	0	1	0	0
4C	26	12	32	17	0.4250	0.4706	0.9942	0.5313	26	35	0	0	2	2
4B	68	10	100	27	0.6750	0.6765	0.9719	0.2700	58	73	5	9	4	4
4A	206	4	306	31	0.7750	0.7353	0.8941	0.1013	112	128	42	74	6	6
3	401	4	707	35[c]	0.8750	0.8529	0.7412	0.0495	40	49	17	22	3	5
2	98	0	805	35	0.8750	0.8529	0.7035	0.0435	3	3	1	1	0	0
1[d]	1832[e]	5	2637	40	1.0000	1.0000	0.0000	0.0152	22	23	6	6	1	1

...hylactic mastectomy which showed DCIS in a patient with contralateral invasive ductal cancer; does not include two negative double prophylactic mastectomies nor 7 other single prophylactic mastectomies.

...s atypical ductal (ADH) or lobular hyperplasia (ALH), LCIS, atypical papilloma, and radial sclerosing lesion

...ive ductal cancer was considered BI-RADS 4a on M, missed on US, and classified as BI-RADS 3 on integrated M+US; this was diagnosed when the participant presented with palpable metastatic adenopathy 264 days after study entry.

...ant with cancer due to melanoma presenting with axillary metastases was classified as BI-RADS 2 on M and BI-RADS 1 on integrated M+US.

...re not asked to integrate findings on M+US if both exams recommended routine follow up: these cases were analyzed as "negative" on M+US. Of 1832 cases analyzed as BI-RADS 1 on M+US, 342 were BI-RADS 1 on M+US, 318 were BI-RADS
...e BI-RADS 2 on both; 2 were BI-RADS 3 on initial mammography and downgraded to BI-RADS 1 after integration with US; the remaining 423 were BI-RADS 1 on both M and US.

Table 4

Summary of Recommendations After Screening Mammography (M), Screening US, Combined M+US, and After Diagnostic Workup

Recommendation	Mammography		US		M+US		After Additional Evaluation	
	Total	Cancers (%)	Total	Cancers (%)	Total	Cancers (%)	Total	Cancers (%)
Additional Imaging	249	15[a] (6.0)	98	4[b] (4.1)	175	8 (4.6)	6	na
Biopsy	21	4 (19.0)	195	12 (6.2)	239	20 (8.4)	272	31 (11.4)
Additional Imaging and Biopsy	6	2 (33.3)	44	6 (13.6)	22	4 (18.2)	4	na
Short-interval follow-up	59	na	227	3 (1.3)	286	3 (1.0)	336	3 (0.9)
Annual follow-up[c]	2302	19 (0.8)	2072	15 (0.7)	1915	5 (0.3)	2019	6 (0.3)
Missing	na	na	1	na	na	na	na	na
Total	2637	40 (1.5)	2637	40 (1.5)	2637	40	2637	40 (1.5)

[a] One participant with a 4 mm invasive ductal cancer was considered BI-RADS 4a on M and recommended for additional imaging, missed on US, and classified as BI-RADS 3 with recommendation for short interval follow-up on integrated M+US; cancer was diagnosed when the participant presented with palpable metastatic adenopathy 264 days after study entry. Another participant was recalled for additional imaging of calcifications felt to be BI-RADS 3 on screening, and benign on additional imaging, with cancer diagnosed 363 days after study entry due to a finding on the (early) 12-month ultrasound screen.

[b] Two participants were classified as BI-RADS 3 on US, but with recommendation for additional imaging. One had a BI-RADS 4b lesion on mammography and was biopsied. The other was recommended for short-interval follow-up after integrated M+US and was diagnosed at 6 month follow-up US.

[c] The participant with cancer due to axillary metastasis from melanoma was recommended for routine follow-up on all imaging.

Table 5

Summary of Cancers Identified by Participant as a Function of Breast Density, Mammogram Type, and Eligibility Risk Factor

	Counts (% of Analysis Set)	All Cancers (# invasive)	# Cancers BIRADS>3 only on Mammography (M)[a] (# invasive)	# Cancers BI-RADS>3 Only on US (#invasive)	#Cancers BIRADS>3 on Both M and US (# invasive)	# Cancers BI-RADS ≤3 on Both M and US (#invasive)[c]
Total	2637 (100.0%)	40 (34)	12 (7)	12 (11)	8 (8)	8 (8)
Breast Density[b]						
≤ 25%	53 (2.0%)	.	.	.	.	.
26–40%	297 (11.3%)	6 (4)	3 (2)	3 (2)	.	.
41–60%	811 (30.7%)	14 (13)	4 (3)	4 (4)	3 (3)	3 (3)
61–80%	968 (36.7%)	11 (9)	3 (1)	2 (2)	4 (4)	2 (2)
>80%	506 (19.2%)	9 (8)	2 (1)	3 (3)	1 (1)	3[c] (3)
Missing	2 (0.1%)	.	.	.	.	.
Total	2637 (100.0%)	40 (34)	12 (7)	12 (11)	8 (8)	8 (8)
Mammogram Type						
Digital	923 (35.0%)	14 (14)	3 (3)	4 (4)	3 (3)	4 (4)
Film-Screen	1714 (64.9%)	26 (20)	9 (4)	8 (7)	5 (5)	4[c] (4)
Total	2637 (100.0%)	40 (34)	12 (7)	12 (11)	8 (8)	8 (8)
Eligibility Risk Factor						
ADH/ALH/LCIS or Atypical Papilloma	83 (3.1%)	1 (.)	1 (.)	.	.	.
Five Year Risk, Gail Model ≥ 2.5%	403 (15.3%)	4 (4)	.	2 (2)	1 (1)	1 (1)
Five Year Risk, Gail Model ≥ 1.7% and Extremely Dense Breasts	223 (8.5%)	2 (1)	1 (.)	.	.	1 (1)
History of Prior Chest and/or Mediastinal and/or Axillary Irradiation	8 (0.3%)	.	.	.	.	.
Lifetime Risk ≥ 25%	497 (18.8%)	5 (4)	3 (2)	1 (1)	.	1 (1)
Mutation in BRCA-1 or BRCA-2	23 (0.9%)	.	.	.	.	.
Personal History of Breast Cancer	1400 (53.1%)	28 (25)	7 (5)	9 (8)	7 (7)	5[c] (5)
Total	2637 (100.0%)	40 (34)	12 (7)	12 (11)	8 (8)	8 (8)

[a] One invasive ductal cancer seen on initial mammography was classified as probably benign after integrated interpretation with US and had breast Density=41–60%, lifetime risk ≥25% and mammogram type film-screen.

[b] Breast density is overall visually estimated breast density. All participants were visually judged to have at least heterogeneously dense parenchyma in at least one quadrant of the breast(s), except for 43 participants who had no prior mammography.

[c] Melanoma case had breast density >80%, personal history of breast cancer, and mammogram type film-screen