Performance of Screening Ultrasound as an Adjunct to Screening Mammography in Women Across the Spectrum of Breast Cancer Risk

Janie M. Lee, MD, MSc[1]
Robert F. Arao, MPH[2,3]
Brian L. Sprague, PhD[4]
Karla Kerlikowske, MD[5]
Constance D. Lehman, MD, PhD[6]
Robert A. Smith, PhD[7]
Louise M. Henderson, PhD[8]
Garth H. Rauscher, PhD[9]
Diana L. Miglioretti, PhD[10,2]

[1]Department of Radiology, University of Washington, and Seattle Cancer Care Alliance, Seattle WA
[2]Kaiser Permanente Washington Health Research Institute, Seattle, WA
[3]Department of Psychiatry and Behavioral Science, University of Washington, Seattle WA
[4]Department of Surgery and Office of Health Promotion Research, College of Medicine, University of Vermont, Burlington, VT
[5]Departments of Medicine and Epidemiology and Biostatistics, General Internal Medicine Section, Department of Veterans Affairs, University of California, San Francisco, San Francisco, CA
[6]Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA
[7]Cancer Control Department, American Cancer Society, Atlanta, GA
[8]Department of Radiology, University of North Carolina, Chapel Hill, Chapel Hill, NC
[9]Department of Epidemiology, School of Public Health, University of Illinois at Chicago, Chicago, IL
[10]Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA

Corresponding Author: Janie M. Lee, MD, MSc
Seattle Cancer Care Alliance
825 Eastlake Avenue East
Seattle, WA 980109
Phone: (206) 606-6777
Fax: (206) 606-6556
Email: jmlee58@uw.edu

Word Count: 2,999 words

Key Points

Question: Does adding whole breast ultrasound to mammography improve breast cancer screening effectiveness?

Findings: For women whose breast cancer risk ranged from low to very high, we observed significantly higher short interval follow-up and biopsy recommendation rates with screening mammography plus same day ultrasound compared to mammography alone, but no significant increase in cancer detection or decrease in interval cancer rates.

Meaning: These results suggest that the benefits of supplemental ultrasound screening may not outweigh associated harms.

ABSTRACT

Importance: Whole-breast ultrasound has been advocated to supplement screening mammography to improve outcomes in women with dense breasts.

Objective: To determine the performance of screening mammography plus screening ultrasound compared with screening mammography alone in community practice.

Design: Observational cohort study.

Setting: The study was IRB approved and HIPAA compliant. Two Breast Cancer Surveillance Consortium (BCSC) registries provided prospectively collected data on screening mammography with vs. without same-day breast ultrasound from 2000-2013.

Participants: 6,081 screening mammography plus same day screening ultrasound examinations in 3,385 women were propensity score matched 1:5 to 30,062 screening mammograms without screening ultrasound in 15,176 women from a sample of 113,293 mammograms. Exclusion criteria included personal history of breast cancer and self-reported breast symptoms.

Exposure: Screening mammography with versus without screening ultrasound.

Main Outcomes and Measures: Cancer detection rate (CDR), and rates of interval cancer, false-positive (FP) biopsy recommendation, short-interval follow-up (SIFU), and positive predictive value of biopsy recommendation (PPV_2) were estimated and compared using log-binomial regression.

Results: Screening mammography with vs without ultrasound examinations were performed more often in women with dense breasts (74% vs 36% in the overall sample); who were younger than 50 years (50% vs 32%), with a family history of breast cancer (43% vs 15%).

While 21% of screening ultrasound examinations were performed in women with high or very high (≥2.50%) BCSC 5-year risk scores, 53% had low or average (<1.67%) risk. Comparing mammography plus ultrasound to mammography alone, CDR was similar: 5.4 vs. 5.5 per 1000 examinations (adjusted relative risk [RR]=1.14 95% confidence interval [CI]: 0.76-1.68); as were interval cancer rates: 1.5 vs. 1.9 per 1,000 examinations (RR=0.67, 95%CI: 0.33-1.37); FP biopsy rates were significantly higher: 52.0 vs. 22.2 per 1000 examinations (RR=2.23, 95%CI: 1.93-2.58); as was SIFU: 3.9% vs. 1.1% (RR=3.10, 95%CI: 2.60-3.70); PPV_2 was significantly lower: 9.5% vs. 21.4% (RR=0.50 95%CI: 0.35-0.71).

Conclusions and Relevance: In a relatively young population of women at low, intermediate, and high breast cancer risk, our results suggest that the benefits of supplemental ultrasound screening may not outweigh associated harms.

INTRODUCTION

Increasing awareness that breast density is a risk factor for developing breast cancer and also makes breast cancer more difficult to detect on mammography has led to grassroots efforts to educate women about breast density. In 2009, Connecticut was the first state to pass legislation requiring that all women receiving mammography be directly informed about breast density and that payors cover supplemental ultrasound screening in women with dense breasts[1]. Since then, at least 34 additional states have enacted breast density notification legislation,[2] and a federal bill has been introduced[3,4]. Seven of these states mandate insurance coverage for screening ultrasound in women with dense breasts. The laws vary across states, but most require notification if a woman's mammographic density is either heterogeneously or extremely dense, as determined by a radiologist according to the American College of Radiology's Breast Imaging Reporting and Data System (BI-RADS)[5,6]. In addition, notification requirements may include statements that women are at higher breast cancer risk because of their breast density, that breast density may adversely affect the ability of mammography to detect breast cancers, and that women with mammographically dense breasts may wish to consider supplemental screening[1,7-9]. Data from the Breast Cancer Surveillance Consortium (BCSC) indicate that 43% of women undergoing screening mammography aged 40-74 have dense breasts, including 57% of women aged 40-44 years[10].

Studies of screening ultrasound have included women with additional risk factors beyond breast density, who were at intermediate to high breast cancer risk either due to a personal history of breast cancer or high-risk benign breast lesions, or because of genetic susceptibility. In addition, most studies of ultrasound screening performance have been conducted in academic medical centers[11]. A recent systematic review of supplemental

screening ultrasound, magnetic resonance imaging (MRI), and digital breast tomosynthesis for women with dense breasts[8] noted that good-quality evidence was sparse, and effects of supplemental screening on breast cancer outcomes remain unclear.

Accurate information on the effectiveness of screening ultrasound is needed to provide guidance on whether widespread use of screening breast ultrasound with screening mammography would be a beneficial screening strategy. We conducted a retrospective study at two BCSC breast imaging registries to assess utilization of screening ultrasound in community practice and to determine the performance of screening mammography plus ultrasound compared to screening mammography alone in women across the spectrum of breast cancer risk.

Methods

Study Setting and Data Sources

We included screening mammograms with or without screening ultrasound performed at breast imaging facilities in one of two BCSC registries[12] (http://www.bcsc-research.org): San Francisco Mammography Registry and Vermont Breast Cancer Surveillance System, which linked woman-level risk factors and clinical information to information on breast imaging examinations with pathology databases, regional Surveillance, Epidemiology, and End Results (SEER) programs, and state tumor registries. BCSC registries and its Statistical Coordinating Center (SCC) received Institutional Review Board approval for active or passive consenting processes or a waiver of consent to enroll participants, link and pool data, and perform analysis. All procedures were Health Insurance Portability and Accountability Act compliant, and registries and the SCC received a Federal Certificate of Confidentiality and other protections for the identities of women, physicians, and facilities.

We identified breast ultrasound examinations with an indication of screening, performed on the same day as a screening mammogram from 2000-2013. Exclusion criteria included: i) personal history of breast cancer, mastectomy, or BI-RADS 6 (known malignancy) assessment, ii) unilateral exam, and iii) self-reported symptoms (except pain). Data abstractors reviewed radiology reports to confirm the ultrasound screening indication. For one registry, after abstracting 13% of reports, 96% were confirmed with screening indication. Thus, we assumed the remaining examinations were performed for screening. For the other registry, all reports were abstracted, and 78% were confirmed with screening indication. Reports with indeterminate indication were reviewed by two radiologists (JML and CDL) with consensus determination of indication. Screening mammograms eligible to be in the matched group were performed at the same facilities, applying the same exclusion criteria as above.

Measures and Definitions

Women completed a questionnaire at each examination to collect information on race and ethnicity, history of first-degree relatives (mother, sister, or daughter) with breast cancer, menopausal status[13], and history of breast biopsy. Prior diagnoses of benign breast disease were collected from pathology databases and grouped into one of four categories: non-proliferative, proliferative without atypia, proliferative with atypia, and lobular carcinoma in situ.[14] American College of Radiology BI-RADS breast density[5,6] was recorded clinically by the interpreting radiologist. The BCSC (version 2.0) 5-year breast cancer risk score was calculated[14].

Performance measures and definitions are provided in Table 1. Because most women with screening ultrasound-detected abnormalities received same day additional imaging, and a

single screening report was issued regardless of whether imaging included only screening or both screening and diagnostic views, recall rate was based on the end-of-day BI-RADS assessment after any additional work-up, and defined as recall for additional imaging that was performed on a different day. A BI-RADS end-of-day assessment of 1 (Negative) or 2 (Benign) was considered negative, and 0 (Needs Additional Evaluation), 3 (Probably Benign), 4 (Suspicious) or 5 (Highly Suspicious) was positive.

All other performance metrics were based on the final assessment[15], which differed from end-of-day assessment only for BI-RADS 0 examinations, which were followed for up to 90 days for the first non-zero BI-RADS assessment. Final assessment of 4 or 5 were considered positive and assessments of 1, 2, or 3 were considered negative. Examinations that could not be resolved to a non-zero assessment [N=5 (0.1%) mammography plus ultrasound, N=104 (0.4%) mammography alone] were excluded from calculations of performance metrics using the final assessment.

For each screening examination, women were followed for 12 months after or until the next screening examination, whichever occurred earlier, for breast cancer diagnoses (either invasive adenocarcinoma or ductal carcinoma in situ [DCIS]). True-positive screens (i.e., screen-detected cancers) were defined as positive screens with a breast cancer diagnosis. False-positive screens were defined as positive screens without a breast cancer diagnosis. Negative screens were defined as true negative if no cancer was diagnosed and false negative if cancer was diagnosed during the follow-up period. Recall rate, biopsy recommendation rate, and cancer detection rate were also compared between the first mammography plus ultrasound screening examination in the BCSC and subsequent examinations. We also estimated the following breast cancer outcomes: percentage of minimal cancer (defined as

DCIS or invasive carcinoma ≤10 mm), percentage of node-negative invasive cancers, percentage of stage 0 and 1 cancers, and median size of invasive cancer.

Statistical Analysis

We used logistic regression to estimate the propensity scores, i.e., the probability of screening with mammography plus ultrasound versus mammography alone, based on: BCSC registry, age [linear and quadratic] and year of examination, time since prior mammogram, race/ethnicity, breast density, menopausal status, family history of breast cancer, and prior benign biopsy result. A SAS macro[16] was used for 1:5 matching of mammography plus ultrasound examinations (n=6,081) to mammography alone (n=30,062) within the same registries and without replacement, using the logit of the propensity score with a caliper width of 0.3 standard deviations. Matching was performed separately for each subgroup needed for each performance measure: all screening examinations for rates per 1000 screens (n=36,143); screening examinations with cancer for sensitivity (n=252); screening examinations without cancer for specificity (n=35,878); and screening exams with positive final assessment for PPV_2 (n=2,062). We compared the covariate distributions in the mammography plus group and the mammography alone group before vs. after matching using the standardized differences of the proportions of each covariate category[17].

We assessed the joint distributions of breast density and BCSC 5-yr risk in the women receiving mammography plus ultrasound screening. Unadjusted performance measures were calculated with 95% confidence intervals (CIs) for the matched groups. We used log-binomial regression to estimate relative risks comparing performance metrics for mammography plus ultrasound vs. matched screens with mammography alone, including the matched set as a random effect to account for correlation among these examinations and adjusting for

characteristics included in the propensity score model to account for potential residual confounding.

Analyses were performed in SAS® software, Version 9.4 (SAS Institute, Cary, NC), and figures were produced using Python 3.4.

RESULTS

We identified 6,081 mammography plus ultrasound examinations in 3,386 women (Table 2). Compared to women in the overall mammography alone group before matching (N=113,293), women receiving mammography plus ultrasound were more likely to be white, non-Hispanic (79% vs. 76%), younger than 50 years old (50% vs. 32%), to have a family history (43% vs. 15%), dense breasts (74% vs. 36%), and high BCSC risk scores (>2.5%, 21% vs. 7%). Of note, 26% of women receiving mammography plus ultrasound did not have dense breasts.

Mammography plus ultrasound examinations were matched 1:5 to 30,062 mammography examinations in 15,176 women (Table 2). Before matching, the differences between the covariate distributions in the overall sample were medium to large[18], with the largest absolute standardized differences for scattered breast density (0.70), no prior biopsy (0.66), and family history (0.65). After matching, absolute standardized differences were small[18], with the largest differences for no prior biopsy (0.25) and family history (0.29). However, after matching, some medium-sized absolute standardized differences remained for the samples used for sensitivity (max=0.50 for no prior biopsy) and PPV (0.46 for family history and 0.43 for no prior biopsy), but differences for specificity sample were small (<0.29) (data not shown). The distribution of propensity scores for mammography plus ultrasound and

mammography alone subgroups demonstrated improved overlap after matching (eFigure 1; sensitivity, specificity, and PPV2, eFigures 2a-c). However, differences remained for: age, menopausal status, family history of breast cancer, examination year, breast density, benign biopsy result, and BCSC 5-year risk. Therefore, we additionally adjusted for these characteristics when comparing performance measures.

Figure 1 presents the joint distribution of BCSC 5-year risk of developing invasive breast cancer and BI-RADS density categories for 5,392 mammography plus ultrasound examinations. While 75% of examinations were performed in women with dense breasts (n=4042, eTable 1), only 21% of these examinations were in women with high or very high 5-year risk (n=1154). Very few women with high or very high risk had non-dense breasts (4%, 152/1154).

Performance of combined mammography and ultrasound screening

Compared with mammography alone (Table 3), mammography plus ultrasound screening was associated with fewer end-of-day assessments for additional imaging (BI-RADS 0: 0.4% versus 17.2%) and lower overall recall rate for additional imaging or biopsy (BI-RADS 0, 3, 4, or 5: 9.9% versus 17.6%, RR= 0.52, 95%CI: 0.48-0.57), indicating that women were less likely to need a second visit to complete diagnostic evaluations. However, with mammography plus ultrasound, the efficiency of same day imaging evaluation was offset by an almost doubling of the biopsy recommendation rate. (57.4 vs. 27.7 per 1000 examinations; RR=2.05, 95%CI: 1.79-2.34). The short interval follow-up rate for probably benign findings was also significantly increased with mammography plus ultrasound (3.9% vs. 1.1%, RR=3.10, 95%CI: 2.60-3.70). In addition, the false-positive biopsy recommendation rate more than

doubled (52.0 vs 22.2 per 1000 examinations, RR=2.23, 95%:CI: 1.93-2.58), with a corresponding decrease in PPV_2 of approximately half (PPV_2 9.5% vs 21.4%, RR=0.50, 95%CI: 0.35-0.71). Increased sensitivity (78.6% vs. 73.8%, RR=1.08, 95%CI: 0.92-1.27) and decreased false-negative rate (1.5 vs 1.9 per 1000 screens, RR=0.67, 95%CI: 0.33-1.37) were observed with mammography plus ultrasound, but these differences were not statistically significant. The cancer detection rate was comparable (5.4 vs. 5.5 per 1000 screens, RR=1.14, 95%CI: 0.76-1.68) between groups.

When recall rate, biopsy recommendation rate, and cancer detection rate were stratified by first versus subsequent mammography plus ultrasound examinations (n=2,040 and 4,041 respectively), all rates declined significantly ($p<0.009$) on subsequent examinations. Recall rate decreased from 12.5% to 6.5% (9.9% overall); biopsy recommendation rate declined from 75 to 49 per 1000 examinations (57 per 1000 overall), and cancer detection rate declined from 10.8 to 2.7 per 1000 examinations (5.4 per 1000 overall).

Breast Cancer Characteristics

There were 42 breast cancers among women with mammography plus ultrasound cohort and 221 in the matched sample (Table 4). Regardless of screening strategy, most breast cancers were invasive ductal carcinomas, small (<20mm in size), ER/PR-positive, and node-negative. Women with mammography plus ultrasound had a higher proportion of stage 0 non-invasive ductal carcinoma in situ: 47.6% (20 of 42, 95%CI: 32.5-62.7%) vs 34.8% in matched controls (77 of 221, 95%CI: 28.6-41.1%), but this difference was not statistically significant (p=0.12).

DISCUSSION

As the number of states with breast density notification laws continues to increase[2], post-legislation reports indicate small but significantly increased use of supplemental ultrasound[7,19,20] and MRI[19,21] among women with mammographically dense breasts, with a greater observed increase in ultrasound utilization compared with MRI. A single report of post-legislation outcomes was conducted in Connecticut[22], where analysis of breast cancers recorded in the Connecticut SEER registry observed a small increase in detection of localized invasive breast cancer, but no association with changes in rates of regional or metastatic stages of disease, compared with control states without breast density legislation. With increasing utilization of supplemental ultrasound screening, it is critical that its impact on outcomes be evaluated.

Our study found that for every 1000 women screened with mammography plus ultrasound, approximately 5 women would be diagnosed with breast cancer, while 57 women would receive a recommendation for biopsy, and 52 of these women would have benign, false-positive results at pathology. An additional 39 women would receive recommendations for short interval imaging follow-up of detected findings, and approximately 2 women would be diagnosed with breast cancer within a year of having negative screening results. Our results of increased biopsy and false-positive biopsy rates, decreased PPV and specificity of supplemental screening ultrasound are consistent with findings from multiple studies conducted in the United States, Europe, and Asia[8,23-28]. In a recent meta-analysis of screening ultrasound studies in women with dense breasts[27], recommendations for further assessment after the addition of ultrasound to mammography screening approximately doubled, and biopsy recommendations rates increased two- to threefold.

Where most other studies have found significant increases in incremental cancer detection rate with mammography plus ultrasound compared with mammography alone[8,27] and with meta-analysis estimating incremental cancer detection at 3.8 per 1000 examinations[27], our study using propensity-score matching and direct adjustment of confounders found comparable cancer detection rates between the two strategies, and a non-significant reduction in interval cancer rates with mammography plus ultrasound screening.

However, incremental cancer detection rates should also be considered in the context of cancer detection rates with mammography alone. The two largest studies of screening ultrasound published to date include a randomized clinical trial (Japan Strategic Anti-Cancer Randomized Trial, J-START[26]) of 72,998 women (36,139 women in the mammography plus ultrasound arm) and a report from an Austrian population-based screening program[24] (66,680 women overall; 31,918 women with dense breasts). In the J-START study, the cancer detection rate in the mammography arm was 3.3 per 1000 screens, and 3.9 per 1000 screens in the mammography plus ultrasound arm (increase of 0.6 per 1000 screens). In the Austrian study, the cancer detection rate with mammography alone was 3.5 per 1000 screens, which increased to 4.0 per 1000 screens when ultrasound was added. For the subgroup of women with dense breasts, the cancer detection rate with mammography alone was 1.8 per 1000 screens, which increased to 2.4 per 1000 screens when ultrasound was added. In our study, cancer detection with mammography alone was 5.5 per 1000 screens, and 5.4 per 1000 screens with mammography plus ultrasound, with no significant difference detected between the two propensity-score matched cohorts.

In the Austrian study with lowest cancer detection in women with dense breasts, the addition of ultrasound provided the largest increase in sensitivity of 19%, from 62% with

mammography alone to 81% with mammography plus ultrasound. In the J-START study, sensitivity in the mammography alone arm was 77%, compared with 91% in the mammography plus ultrasound arm (increase of 13%). Our study demonstrated the smallest, and non-significant, increase in sensitivity (6%) from 74% with mammography alone and 79% with mammography plus ultrasound, with substantial overlap of 95% confidence intervals. The Austrian study did not report specificity values. When comparing sensitivity and specificity together, the diagnostic test performance of mammography plus ultrasound in J-START was higher (sensitivity 91% and specificity of 88%) compared with our study (sensitivity 79% and specificity of 95%). The differences in test performance across studies may reflect differences across study populations, the difference between study settings (randomized trial in Japan vs. population-based screening in Austria vs. community-based screening in the United States) and different interpretive thresholds across countries.

In the Austrian study, since all women received mammography first, and then supplemental ultrasound, interval cancer rates could not be compared across strategies. The overall interval cancer rate was 0.4 per 1000 screens. Similar to the J-START study, our study reported reduction in false-negative rates, from 1.9 per 1000 screens with mammography alone to 1.5 per 1000 screens with mammography plus ultrasound (absolute difference of 0.4 per 1000 screens). This difference was not significant given our smaller breast cancer sample of 42 cancers, but the 0.5 per 1000 reduction in false-negative cancer rate in the J-START study (from 1.0 per 1000 screens with mammography alone to 0.5 per 1000 screens with mammography plus ultrasound) was statistically significant.

The results presented in this study reflect real-world clinical practice in the United States for women across the spectrum of breast cancer risk who received same-day, supplemental

ultrasound screening, adding information about the incremental performance and outcomes of supplemental screening with ultrasound compared with mammography alone. Of note, we observed that supplemental ultrasound screening was utilized not only in women with dense breasts; 26% of women receiving it had non-dense breasts. In our sample of breast imaging examinations with comprehensive capture of cancer outcomes for performance assessment, we found no significant screening benefit as measured by increased cancer detection rate, decreased false-negative rate, or increased sensitivity. Our results may reflect the high proportion of women who were at low or average 5-year risk (53.6%) in our study, higher proportion of women with non-dense breasts, lower screening ultrasound sensitivity outside of clinical trial settings, or a combination of these factors. Alternatively, it may reflect the relatively small number of cancers in the mammography plus ultrasound and matched mammography cohorts, with lack of power to detect small differences in measures. Our study had >99% power to detect an incremental cancer detection rate of 3.8 per 1000 as estimated by a recent meta-analysis[27], suggesting that an effect of this magnitude is highly unlikely in the facilities we studied. However, the wide confidence interval around the estimated relative CDR for mammography plus ultrasound versus mammography alone is comparable with values from 0.76 to 1.68. Larger studies are needed for more precise estimates.

Limitations of this study include lack of information on the experience and expertise of the personnel performing the ultrasound examinations. We also did not collect information on whether ultrasound examinations were performed using hand-held ultrasound or automated whole breast ultrasound (ABUS) devices. ABUS devices are thought to increase consistency of image acquisition and decrease operator dependence which limits hand held ultrasound

examinations[29,30]. We also did not fully abstract screening ultrasound reports in one registry, leaving potential for misclassification of 4% of examinations.

The proportions of axillary lymph node-positive breast cancers and false-negative rates were not significantly different for mammography with vs. without ultrasound. A more definitive study, such as a randomized clinical trial conducted in the United States, to evaluate either of these measures, which are thought to correlate with downstream improvement in outcomes for women receiving screening, would require a very large sample size. This is especially true if the primary outcome was reduction in false-negative rate, which would need to be powered to detect a difference of 4 per 10,000 women screened as reported in our study.Currently, the only breast cancer guidelines to include supplemental ultrasound screening are those of the American College of Radiology[31], which support consideration of ultrasound for women with elevated risk who would quality for but cannot undergo breast MRI, and for women with increased breast density, “after weighing benefits and risks”.

In conclusion, our observational study of ultrasound screening in women across a range of breast cancer risk found modest, non-significant benefits and rates of screening harms that were high and consistent with prior reports. To apply supplemental ultrasound screening with greater effectiveness, additional efforts are needed to more accurately identify women who will benefit from supplemental screening, as well as development of the capacity to deliver high-quality supplemental screening, and additional interventions to reduce the frequency of screening-related harms.

ACKNOWLEDGMENTS

Author Contributions: Dr. Lee, Dr. Miglioretti, and Mr. Arao had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of Interest Disclosures: Drs. Lee and Lehman each report research grant funding and a consulting agreement with GE Healthcare. The remaining authors disclose no conflicts of interest.

Funding/Support: This research was funded by the Breast Cancer Surveillance Consortium program project (P01CA154292). Additional support was provided by the American Cancer Society through a gift from the Longaberger Company's Horizon of Hope Campaign (ACS 15922 #NHLONGGBR). Data collection for this work was additionally supported, in part, by funding from the National Cancer Institute (U54CA163303). The collection of cancer and vital status data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S.

For a full description of these sources, please see: **http://www.bcsc-research.org/work/acknowledgement.html**. We thank the participating women, mammography facilities, and radiologists for the data they have provided for this study. You can learn more about the BCSC at: **http://www.bcsc-research.org/**.

Role of the Funder/Sponsor: The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and the decision to submit the manuscript for publication.

REFERENCES

1. Dehkordy SF, Carlos RC. Dense breast legislation in the United States: State of the states. *Journal of the American College of Radiology.* 2013;10:899-902.
2. Are You Dense Advocacy. State density reporting efforts. 2018; https://www.areyoudenseadvocacy.org/dense/. Accessed 02/12/2018.
3. 115th Congress (2017-2018). S.2006 Breast Density and Mammography Reporting Act of 2017, introduced in U.S. Senate 10/25/2017. 2017; https://www.congress.gov/bill/115th-congress/senate-bill/2006/text?q=%7B%22search%22%3A%5B%22Health%2C+Education%2C+Labor%2C+and+Pensions%22%5D%7D&r=2. Accessed 02/18/2018, 2018.
4. 115th Congress (2017-2018). H.R.4122 Breast Density and Mammography Reporting Act of 2017, introduced in House 10/25/2017. 2017; https://www.congress.gov/bill/115th-congress/house-bill/4122/text. Accessed 02/18/2018.
5. American College of Radiology. *American College of Radiology Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas).* 4th edition ed. Reston, VA: Am Coll Radiol; 2003.
6. Radiology ACo. *American College of Radiology Breast Imaging Reporting and Data System Atlas (BI-RADS® Atlas), 5th ed.* Reston, VA: American College of Radiology; 2013.
7. Horny M, Cohen AB, Duszak R, Christiansen CL, Shwartz M, Burgess JF. Dense breast notification laws: Impact on downstream imaging after screening mammography. *Medical Care Research and Review.* 2018;1077558717751941.

8. Melnikow J, Fenton JJ, Whitlock EP, et al. Supplemental screening for women with dense breasts: A systematic review fo the U. S. Preventive Services Task Force. *Annals of Internal Medicine.* 2016;164:268-278.
9. Haas JS, Kaplan CP. The divide between breast density notification laws and evidence-based guidelines for breast cancer screening. . *JAMA Internal Medicine.* 2015;175:1439-1440.
10. Sprague BL, Gagnon RE, Burt V, et al. Prevalence of mammographically dense breasts in the United States. *Journal of the National Cancer Institute.* 2014;106.
11. Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *American Journal of Obstetrics and Gynecology.* 2015;212:9-17.
12. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. *AJR Am J Roentgenol.* 1997;169(4):1001-1008.
13. Phipps AI, Ichikawa L, Bowles EJ, et al. Defining menopausal status in epidemiologic studies: A comparison of multiple approaches and their effects on breast cancer rates. *Maturitas.* 2010;67(1):60-66.
14. Tice JA, Miglioretti DL, Li CS, Vachon CM, Gard CC, Kerlikowske K. Breast Density and Benign Breast Disease: Risk Assessment to Identify Women at High Risk of Breast Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology.* 2015;33(28):3137-3143.

15. Yankaskas BC, Taplin SH, Ichikawa L, et al. Association between mammography timing and measures of screening performance in the United States. *Radiology.* 2005;234:363-373.

16. Coca-Perraillon M. Local and global optimal propensity score matching (Paper 185-2007). *SAS Global Forum 2007*. Cary, NC.: SAS Institute Inc.; 2007.

17. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment gorups in propensity-score matched samples. *Statistics in Medicine.* 2009;28:3083-3107.

18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd edition ed. Hillsdale, NJ.: Lawrence Erlbaum Associates Publishers; 1988.

19. Sanders LA, King AB, Goodman KS. Impact of the New Jersey breast density law on imaging and intervention volumes and breast cancer diagnosis. *Journal of the American College of Radiology.* 2016;13:1189-1194.

20. Sobotka J, Hinrichs C. Breast density legislation: Discussion of patient utilization and subsequent direct financial ramifications for insurance providers. *Journal of the American College of Radiology.* 2015;12:1011-1015.

21. Chau SL, Alabster A, Luikart K, Brenman LM, Habel LA. The effect of California's breast density notification legislation on breast cancer screening. *Journal of Primary Care & Community Health.* 2017;8:55-62.

22. Richman I, Asch SM, Bendavid E, Bhattacharya J, Owens DK. Breast density notification legislation and breast cancer stage at diagnosis: Early evidence from the SEER registry. *Journal of General Internal Medicine.* 2017;32:603-609.

23. 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;299:2151-2163.

24. Buchberger W, Geiger-Gritsch S, Knapp R, Gautsch K, Oberaigner W. Combined screening with mammography and ultrasound in a population-based screening program. *European Journal of Radiology.* 2018;101:24-29.

25. Corsetti V, Houssami N, Ghirardi M, et al. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: Interval breast cancers at 1 year follow up. *European Journal of Cancer.* 2011;47:1021-1026.

26. Ohuchi N, Suzuki A, Sobue T, et al. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomized controlled trial. *Lancet.* 2016;387:341-348.

27. Rebolj M, Assis V, Brentnall A, Parmar D, Duffy SW. Addition of ultrasound to mammography in the case of dense breast tissue: systematic review and meta-analysis. *British Journal of Cancer.* 2018;118:1559-1570.

28. Tagliafico AS, Calabrese M, Mariscotti G, et al. Adjunct screening with tomosynthesis or ultrasound in women with mammography-negative dense breasts: Interi report of a prospective comparative trial. *Journal of Clinical Oncology.* 2016;34:1882-1888.

29. Lander MR, Tabar L. Automated 3-D breast ultrasound as a promising adjunctive screening tool for examinating dense breast tissue. *Seminars in Roentgenology.* 2011;46:302-308.

30. Vourtsis A, Kachulis A. The performance of 3D ABUS versus HHUS in the visualizatin and BI-RADS characterisation of breast lesions in a large cohort of 1,886 women. *European Radiology.* 2018;28:592-601.

31. Monticciolo DL, Newell MS, Moy L, Niell B, Monsees B, Sickles EA. Breast cancer screening in women at higher-than-average risk: Recommendations from the ACR. *Journal of the American College of Radiology.* 2018;15:408-414.

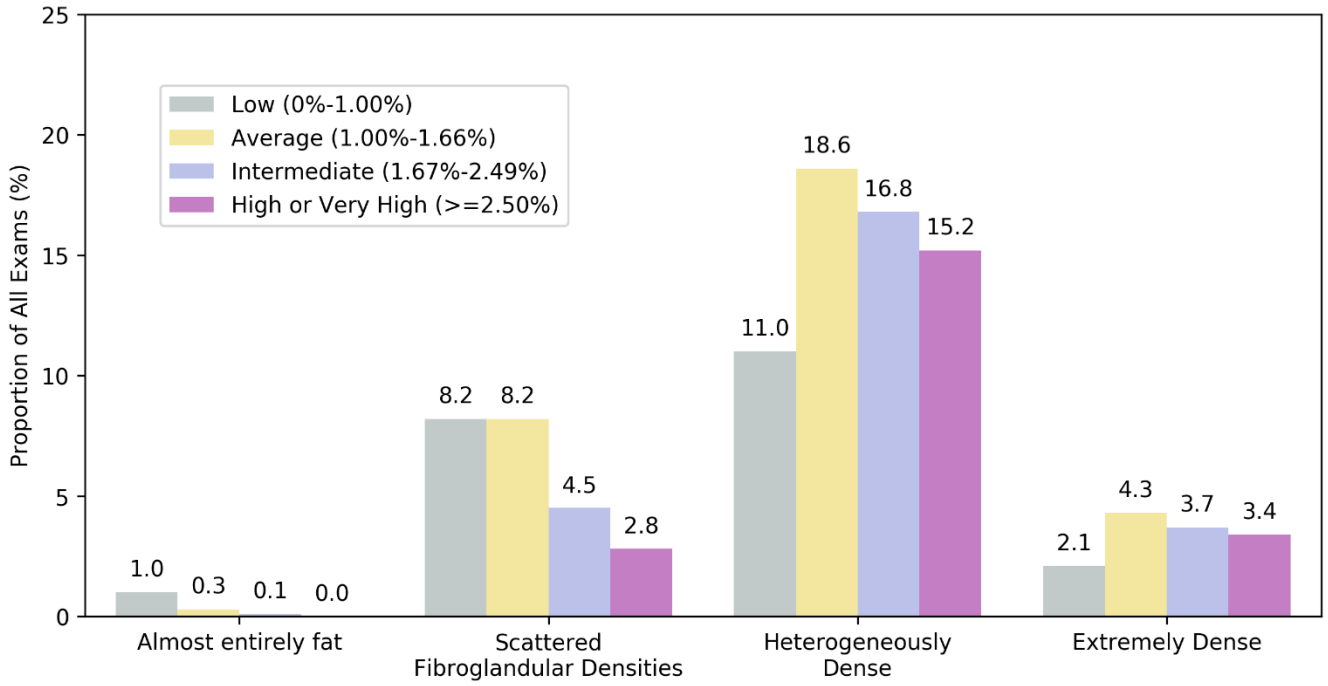

Figure 1. Joint distributions of BCSC 5-yr risk by BI-RADS breast density category in 5,392 combined mammography and ultrasound screening examinations.

Table 1. Performance measures and definitions

Performance Measure	Definition
Recall rate	Percentage of screening exams with a positive end-of-day assessment
Cancer detection rate	Number of true-positive screens per 1000 examinations
Interval cancer rate	Number of screening exams with a negative final assessment and cancer diagnosed within the follow-up period per 1000 screens
Cancer rate	Number of screens with cancer within the follow-up period per 1000 screens
Biopsy recommendation rate	Number of screening exams with a positive final assessment per 1000 screens
False-positive biopsy recommendation rate	Number of screening exams with a false-positive final assessment per 1000 screens
Positive predictive value of biopsy recommendation (PPV_2)	Percentage of category 4 and 5 assessments with a tissue diagnosis of cancer within the follow-up period
Short-interval follow up rate	Percentage of screens with a final category 3 assessment
Sensitivity	Percentage of true positive results among those with cancer within the follow-up period
Specificity	Percentage of true negative results among those without cancer within the follow-up period

Table 2. Characteristics of mammography plus ultrasound examination cohort and the matched sample and total population of women receiving screening mammography alone.

Characteristics	Mammography plus Ultrasound n (%)	Mammography alone (matched) n (%)	Standardized mean differences after matching	Mammography alone (overall) n (%)	Standardized mean differences before matching
Total	6,081	30,062		113,293	
Age, years*					
30-39	506 (8.3)	1,093 (3.6)	0.20	2,800 (2.5)	0.26
40-49	2,516 (41.4)	12,636 (42.0)	-0.01	33,114 (29.2)	0.26
50-59	2,052 (33.7)	10,477 (34.9)	-0.03	32,803 (29.0)	0.10
60-69	745 (12.3)	4,130 (13.7)	-0.04	23,034 (20.3)	-0.22
70-79	212 (3.5)	1,326 (4.4)	-0.05	13,857 (12.2)	-0.33
>=80	50 (0.8)	400 (1.3)	-0.05	7,685 (6.8)	-0.32
Race*					
White, non-Hispanic	2,626 (79.4)	13,330 (80.2)	-0.02	46,107 (76.0)	0.08
Black, non-Hispanic	11 (0.3)	76 (0.5)	-0.03	654 (1.1)	-0.10
Asian/Pacific Islander	396 (12.0)	1,800 (10.8)	0.04	7,681 (12.7)	-0.02
Hispanic	213 (6.4)	1,106 (6.7)	-0.01	4,685 (7.7)	-0.05
Mixed/Other	62 (1.9)	316 (1.9)	0.00	1,534 (2.5)	-0.04
Menopausal status*					
Pre	2,074 (34.1)	11,665 (38.8)	-0.10	30,869 (27.2)	0.15
Post	2,723 (44.8)	12,067 (40.1)	0.10	63,224 (55.8)	-0.22
Surgical/Other amenorrhea/Unknown	1,284 (21.1)	6,330 (21.1)	0.00	19,200 (16.9)	0.11
First degree family history of breast cancer*					
No	3,460 (57.1)	21,227 (71.0)	-0.29	95,565 (85.0)	-0.65
Yes	2,595 (42.9)	8,688 (29.0)	0.29	16,897 (15.0)	0.65
Unknown	26	147		831	
Exam Year*					
2005-2006	285 (4.7)	1,784 (5.9)	-0.05	16,635 (14.7)	-0.34
2007-2008	1,284 (21.1)	6,671 (22.2)	-0.03	31,164 (27.5)	-0.15
2009-2010	1,999 (32.9)	9,894 (32.9)	0.00	32,216 (28.4)	0.10
2011-2013	2,513 (41.3)	11,713 (39.0)	0.05	33,278 (29.4)	0.25
Time since last mammogram*					
None	92 (1.6)	517 (1.8)	-0.02	4,386 (4.1)	-0.15
1-2 years	5,398 (95.6)	26,493 (94.8)	0.04	95,293 (89.2)	0.24
3 or more years	157 (2.8)	948 (3.4)	-0.03	7,123 (6.7)	-0.18

Unknown	434	2,104		6,491	
Breast Density*					
Almost entirely fat	83 (1.4)	563 (1.9)	-0.04	8,226 (7.4)	-0.30
Scattered	1,410 (24.3)	9,138 (31.6)	-0.16	63,152 (56.7)	-0.70
Heterogenously dense	3,543 (61.0)	17,126 (59.2)	0.04	37,074 (33.3)	0.58
Extremely dense	774 (13.3)	2,103 (7.3)	0.20	2,854 (2.6)	0.40
Unknown	271	1,132		1,987	
Benign Biopsy Result*					
No prior biopsy	3,045 (50.1)	18,786 (62.5)	-0.25	90,745 (80.1)	-0.66
Biopsy, pathology unknown	1,446 (23.8)	6,696 (22.3)	0.04	15,878 (14.0)	0.25
Non-proliferative disease	831 (13.7)	2,956 (9.8)	0.12	4,320 (3.8)	0.36
Proliferative without atypia	620 (10.2)	1,290 (4.3)	0.23	1,808 (1.6)	0.37
Proliferative with atypia	109 (1.8)	287 (1.0)	0.07	418 (0.4)	0.13
Lobular Carcinoma in Situ	30 (0.5)	47 (0.2)	0.05	124 (0.1)	0.07
BCSC 5-yr risk					
Low (0%-1.00%)	1,196 (22.2)	8,146 (29.2)	-0.16	37,108 (38.1)	-0.35
Average (1.00%-1.66%)	1,693 (31.4)	9,573 (34.3)	-0.06	37,693 (38.7)	-0.15
Intermediate (1.67%-2.49%)	1,349 (25.0)	6,317 (22.6)	0.06	16,127 (16.6)	0.21
High (2.50%-3.99%)	976 (18.1)	3,402 (12.2)	0.17	5,760 (5.9)	0.38
Very High (>= 3.99%)	178 (3.3)	493 (1.8)	0.10	705 (0.7)	0.19
Unknown	689	2,131		15,900	

* variables controlled for in calculating propensity score

Table 3. Estimated performance measures and results from log-binomial regression analysis.

	Mammography + Ultrasound	Mammography Alone (matched)	Relative Risk^ (95% CI)
End-of-day assessment, N (%)			
1 Negative	2,689 (44.2)	23,909 (79.5)	
2 Benign	2,793 (45.9)	878 (2.9)	
3 Probably Benign	234 (3.9)	64 (0.2)	
0 Needs Additional Imaging	21 (0.4)	5,159 (17.2)	
4 Suspicious	342 (5.6)	49 (0.2)	
5 Highly Suspicious	2 (0.0)	3 (0.0)	
Total	6,081 (100%)	30,062 (100%)	
Performance based on end-of-day assessment			
Recall rate for additional imaging (95% CI)	9.9% (9.1, 10.6)	17.6% (17.1, 18.0)	**0.52** (0.48, 0.57)
End-of-day Assessment of 0,3,4, 5, N	599	5,275	
Total examinations, N	6,081	30,062	
Final assessment, N (%)			
1 Negative	2,694 (44.3)	26,848 (89.3)	
2 Benign	2,798 (46.0)	1,936 (6.4)	
3 Probably Benign	235 (3.9)	341 (1.1)	
0 Needs Additional Imaging	5 (0.1)	104 (0.4)	
4 Suspicious	347 (5.7)	792 (2.6)	
5 Highly Suspicious	2 (0.0)	41 (0.1)	
Total	6,081 (100%)	30,062 (100%)	
Performance based on final assessment			
Biopsy recommendation rate per 1000 (95% CI)	57.4 (51.9, 63.5)	27.7 (25.9,29.7)	**2.05** (1.79, 2.34)
Final Assessment of (4, 5)*, N	349	833	
Total examinations, N	6,081	30,062	
Short Interval Follow-up rate (95%CI)	3.9% (3.4, 4.4)	1.1% (1.0, 1.3)	**3.10** (2.60, 3.70)
Final Assessment of 3, N	235	341	
Total examinations, N	6,081	30,062	

Sensitivity (95% CI)	78.6% (67.1, 92.0)	73.8% (68.1, 80.0)	1.08 (0.92, 1.27)
Final Assessment of 4,5 and cancer, N	33	155	
Total Cancers, N	42	210	
Specificity (95% CI)	94.8% (94.2, 95.3)	97.7% (97.6, 97.9)	**0.97** (0.97, 0.98)
Final Assessment of 0,1,2,3 and no cancer, N	5,719	29,169	
Non-cancers, N	6,035	29,843	
PPV2 (95% CI)	9.5% (6.8, 13.1)	21.4% (19.6, 23.5)	**0.50** (0.35, 0.71)
Final Assessment of 4,5 and cancer	33	367	
Biopsy recommended (Final Assessment of 4,5)	349	1,713	
CDR per 1,000 (95% CI)	5.4 (3.9, 7.6)	5.5 (4.7, 6.4)	1.14 (0.76, 1.68)
Final Assessment of 4,5 and cancer	33	165	
Total examinations	6,081	30,062	
False Negative Rate per 1000 (95% CI)**	1.5 (0.8, 2.8)	1.9 (1.4, 2.4)	0.67 (0.33,1.37)
Final Assessment of 0,1,2,3 and cancer, N	9	56	
Total examinations, N	6,081	30,062	
Cancer Rate per 1,000 (95% CI)	6.9 (5.1, 9.3)	7.4 (6.4, 8.4)	0.99 (0.70, 1.42)
All cancers, N	42	221	
Total examinations, N	6,081	30,062	
FP biopsy recommendation rate per 1000 (95% CI)	52.0 (46.7, 57.8)	22.2 (20.6,24.0)	**2.23** (1.93, 2.58)
Final Assessment of (4, 5) and no cancer, N	316	668	
Total examinations, N	6,081	30,062	

^ Relative risk is from log binomial model adjusted for site, age, examination year, history of biopsy, family history, menopause status, and correlation among women within the same matched set using generalized estimated equations. Sensitivity was adjusted for site and history of breast biopsy

*Final Assessment of (4,5) considered a positive examination result and used to calculate performance measures

** FN rate includes both invasive and DCIS

CI-confidence interval

Table 4. Characteristics of breast cancers occurring within one year of the screening exam for women receiving mammography plus ultrasound and the matched sample and total population of women receiving screening mammography alone

Characteristics	Mammography + Ultrasound	Mammography alone (matched)	Mammography alone (overall)
Cancer Histology			
Non-invasive (DCIS)	20 (47.6)	77 (34.8)	249 (34.6)
Invasive	22 (52.4)	144 (65.2)	470 (65.4)
Ductal	19 (95.0)	113 (80.1)	384 (83.8)
Lobular	1 (5.0)	20 (14.2)	50 (10.9)
Mixed	0 (0.0)	8 (5.7)	24 (5.2)
Other/Unknown	2	3	12
Invasive tumor size			
1-5 mm	2 (10.0)	8 (5.7)	48 (10.6)
6-10 mm	5 (25.0)	37 (26.4)	116 (25.6)
11-15 mm	3 (15.0)	25 (17.9)	87 (19.2)
16-20 mm	3 (15.0)	26 (18.6)	74 (16.3)
> 20 mm	7 (35.0)	44 (31.4)	129 (28.4)
Unknown	2	4	16
Minimal Cancer			
No	13 (31.7)	95 (45.2)	290 (42.0)
Yes	28 (68.3)	115 (54.8)	401 (58.0)
Unknown	1	11	28
Axillary lymph node status			
Negative	35 (85.4)	183 (83.6)	620 (87.6)
Positive	6 (14.6)	36 (16.4)	88 (12.4)
Unknown	1	2	11
AJCC Cancer stage			
0	20 (48.8)	77 (35.2)	249 (35.3)
I	11 (26.8)	85 (38.8)	297 (42.1)
II	8 (19.5)	43 (19.6)	131 (18.6)
III	2 (4.9)	13 (5.9)	24 (3.4)
IV	0 (0.0)	1 (0.5)	5 (0.7)
Unknown	1	2	13
Grade of invasive cancer			

Grade 1	8 (42.1)	48 (34.8)	173 (38.6)
Grade 2	8 (42.1)	62 (44.9)	191 (42.6)
Grade 3	3 (15.8)	28 (20.3)	84 (18.8)
Unknown	3	6	22
Hormone receptor status of invasive cancer			
ER+ or PR+	17 (81.0)	134 (94.4)	426 (93.2)
ER- and PR-	4 (19.0)	8 (5.6)	31 (6.8)
Unknown	1	2	13

DCIS = ductal carcinoma in situ; AJCC = American Joint Committee on Cancer; ER= estrogen receptor; PR=progesterone receptor

*invasive cancer $\leq$1cm, or DCIS of any size