Annals of Internal Medicine

REVIEW

Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Services Task Force

Joy Melnikow, MD, MPH; Joshua J. Fenton, MD, MPH; Evelyn P. Whitlock, MD, MPH; Diana L. Miglioretti, PhD; Meghan S. Weyrich, MPH; Jamie H. Thompson, MPH; and Kunal Shah

Background: Screening mammography has lower sensitivity and specificity in women with dense breasts, who experience higher breast cancer risk.

Purpose: To perform a systematic review of reproducibility of Breast Imaging Reporting and Data System (BI-RADS) density categorization and test performance and clinical outcomes of supplemental screening with breast ultrasonography, magnetic resonance imaging (MRI), and digital breast tomosynthesis (DBT) in women with dense breasts and negative mammography results.

Data Sources: MEDLINE, PubMed, EMBASE, and Cochrane database from January 2000 to July 2015.

Study Selection: Studies reporting BI-RADS density reproducibility or supplemental screening results for women with dense breasts.

Data Extraction: Quality assessment and abstraction of 24 studies from 7 countries; 6 studies were good-quality.

Data Synthesis: Three good-quality studies reported reproducibility of BI-RADS density; 13% to 19% of women were recategorized between "dense" and "nondense" at subsequent screening. Two good-quality studies reported that sensitivity of ultrasonography for women with negative mammography results ranged from 80% to 83%; specificity, from 86% to 94%; and positive predictive value (PPV), from 3% to 8%. The sensitivity of MRI ranged from 75% to 100%; specificity, from 78% to 94%; and PPV, from 3% to 33% (3 studies). Rates of additional cancer detection with ultrasonography were 4.4 per 1000 examinations (89% to 93% invasive); recall rates were 14%. Use of MRI detected 3.5 to 28.6 additional cancer cases per 1000 examinations (34% to 86% invasive); recall rates were 12% to 24%. Rates of cancer detection with DBT increased by 1.4 to 2.5 per 1000 examinations compared with mammography alone (3 studies). Recall rates ranged from 7% to 11%, compared with 7% to 17% with mammography alone. No studies examined breast cancer outcomes.

Limitations: Good-quality evidence was sparse. Studies were small and CIs were wide. Definitions of recall were absent or inconsistent.

Conclusion: Density ratings may be recategorized on serial screening mammography. Supplemental screening of women with dense breasts finds additional breast cancer but increases false-positive results. Use of DBT may reduce recall rates. Effects of supplemental screening on breast cancer outcomes remain unclear.

Primary Funding Source: Agency for Healthcare Research and Quality.

Ann Intern Med. 2016;164:XXX-XXX. doi:10.7326/M15-1789 **www.annals.org**
For author affiliations, see end of text.
This article was published at www.annals.org on 12 January 2016.

Dense breasts are defined by mammographic appearance. The American College of Radiology's (ACR's) Breast Imaging Reporting and Data System (BI-RADS) classifies breasts as almost entirely fatty (BI-RADS category a), scattered areas of fibroglandular density (category b), heterogeneously dense (category c), or extremely dense (category d).

About 27.6 million (43%) women aged 40 to 74 years in the United States have dense breasts; most of these are classified as category c (1). Higher breast density is associated with decreased mammographic sensitivity and specificity and also with increased breast cancer risk. The relative hazard of breast cancer for women with dense breasts ranged from 1.50 (women aged 65 to 74 years) to 1.83 (women aged 40 to 49 years) in an analysis of 1 169 248 women enrolled in the Breast Cancer Surveillance Consortium (unpublished data). Increased breast density has been associated with hormone replacement therapy use, younger age, and lower body mass index (2). Data on breast density and race or ethnicity are limited. In the United States, Asian women have higher breast density (3) but lower than average incidence of breast cancer (4). Increased breast density is not associated with higher breast cancer mortality among women with dense breasts diagnosed with breast cancer, after adjustment for stage and mode of detection (5).

Supplemental breast cancer screening with additional screening modalities has been proposed to improve the early detection of breast cancers. No clinical guidelines explicitly recommend use of supplemental breast cancer screening on women with dense breasts (6–9), but as of September 2015, 24 states had enacted legislation requiring that women be notified of breast density with their mammography results; 9 more states are considering mandatory notification (10) (**Appendix Table 1**, available at www.annals.org). Most states require specific language distinguishing dense (BI-RADS c and d) from nondense breasts, and 4 states require that insurers cover subsequent examinations and tests for women with dense breasts (11–14). Federal legisla-

See also:

Related article 1
Editorial comment 2

tion requiring breast density notification is pending (15).

This report summarizes a systematic review of current evidence on the reproducibility of BI-RADS breast density determinations and on test performance characteristics and outcomes of supplemental screening of women with dense breasts by using hand-held ultrasonography (HHUS), automated whole-breast ultrasonography (ABUS), breast magnetic resonance imaging (MRI), and digital breast tomosynthesis (DBT). Mandatory reporting laws frame notification of women as dense/nondense, so this review focused on this categorization.

METHODS

The review protocol included an analytic framework with 4 key questions (KQs) (**Appendix Figure 1**, available at www.annals.org). Detailed methods, including search strategies, detailed inclusion criteria, and excluded studies, are available in the full evidence report (16).

Data Sources and Searches

MEDLINE, PubMed, EMBASE, and the Cochrane Library were searched for relevant English-language studies published between January 2000 and July 2015. We reviewed reference lists from retrieved articles and references suggested by experts.

Study Selection

Two investigators independently reviewed abstracts and full-text articles for inclusion according to predetermined criteria (E.P.W. and J.H.T. for KQ 1, J.M. and J.J.F. for KQs 2 to 4). Included studies examining the reproducibility of BI-RADS breast density categorization focused on asymptomatic women aged 40 years or older undergoing digital or film mammography. Included studies on supplemental screening with HHUS, ABUS, MRI, or DBT reported outcomes for asymptomatic women with dense breasts aged 40 years and older. In studies that focused primarily on women at high risk for breast cancer (including those with preexisting breast cancer or high-risk breast lesions [such as ductal carcinoma in situ, atypical hyperplasia, and lobular carcinoma in situ], *BRCA* mutations, familial breast cancer syndromes, or previous chest-wall radiation) and studies that included women with nondense breasts, we analyzed the relevant subset when available in the publication or provided by the authors.

A priori inclusion criteria limited studies on BI-RADS reproducibility to fair- or good-quality randomized, controlled trials; cohort studies; or test sets involving multiple blind readings by at least 3 readers. Studies on test performance characteristics and outcomes of supplemental screening modalities were limited to fair- or good-quality randomized, controlled trials; cohort studies; or diagnostic accuracy studies with reference standards applied to all participants. We examined sensitivity, specificity, positive predictive values (PPVs), negative predictive values (NPVs), and available clinical outcomes (including cancer detection rates, recall rates, and biopsy rates). We defined recall as the need for any additional diagnostic testing after supplemental screening, including imaging and biopsy.

Data Extraction and Quality Assessment

Two investigators (E.P.W. and J.H.T. for KQ 1, J.M. and J.J.F. for KQs 2 to 4) critically appraised all included studies independently using the U.S. Preventive Services Task Force's (USPSTF's) design-specific criteria (17), supplemented with the National Institute for Health and Clinical Excellence methodology checklists (18) and the Quality Appraisal Tool for Studies of Diagnostic Reliability (19). According to USPSTF criteria, a good-quality study generally met all prespecified criteria; fair-quality studies did not meet all criteria but had no important limitations. Poor-quality studies had important limitations that could invalidate results (inadequate or biased application of reference standard; population limited to very high-risk patients).

Data Synthesis and Analysis

When available or provided by the authors, results of supplemental screening for subgroups of women with dense breasts were extracted; we excluded those with other risk factors for breast cancer. We calculated the sensitivity and specificity of the supplemental breast screening tests for women with negative mammography results. Only cancers detected by the supplemental test after negative mammography results and cancers found at interval follow-up were included. Hence, the values reported represent the sensitivity and specificity for detection of additional cancer in women with negative mammography findings. Similarly, we defined cancer detection rates, recall rates, and biopsy rates to include only those cancer cases, recalls, and biopsies related to supplemental screening after negative results on mammography. Meta-analysis was not performed because there were few good-quality studies.

Role of the Funding Source

This research was funded by the Agency for Healthcare Research and Quality (AHRQ) under a contract to support the work of the USPSTF. The investigators worked with USPSTF members to develop and refine the scope, analytic frameworks, and KQs. AHRQ had no role in study selection, quality assessment, synthesis, or development of conclusions. AHRQ provided project oversight; reviewed the draft report; and distributed the draft for peer review, including to representatives of professional societies and federal agencies. AHRQ performed a final review of the manuscript to ensure that the analysis met methodological standards. The investigators are solely responsible for the content and the decision to submit the manuscript for publication.

RESULTS

The literature search yielded 2067 unique citations; 128 full-text articles considered potentially relevant were reviewed to identify 24 unique studies meeting inclusion criteria (**Appendix Figure 2**, available at www.annals.org). **Table 1** (20–43) provides the characteris-

Table 1. Characteristics of Included Studies

Study, Year (Reference) USPSTF Quality Rating	Design	Country	Examinations/ Women Analyzed	Follow-up Period, *mo*	Population Characteristics
Breast Imaging Reporting and Data System density assessment					
Harvey et al, 2013 (20) Good	Cohort	United States	871 502 examinations 435 751 women	<36*	Age: 58.8 y (mean)
Redondo et al, 2012 (21) Good	Stratified random sample	Spain	100 examinations 100 women	6*	Age: 50–64 y (range)
Spayne et al, 2012 (22) Good	Cohort	United States	11 755 women	3–24*	Age: 66 y (median)
Bernardi et al, 2012 (23) Fair	Test set	Italy	100 examinations 100 women	NA	Age: 43 y (median)
Gard et al, 2015 (24) Fair	Test set	United States	341 women	6*	NR
Hand-held ultrasonography					
Berg et al, 2012 (25) Good	Test accuracy	United States	3414 examinations 1216 women	>12	Age: 55.2 y (mean)† Dense breasts: 100% Personal history: 0% Family history: NR
Corsetti et al, 2011 (26) Good	Test accuracy	Italy	7224 examinations 3356 women	12	Age: 55% <50 y Dense breasts: 100% Personal history: NR Family history: NR
Brancato et al, 2007 (27) Fair	Cohort	Italy	5227 women	NR	Age: 68% 40–49 y Dense breasts: 100% Personal history: NR Family history: NR
Girardi et al, 2013 (28) Fair	Cohort	United States	9960 women	12	Age: 51.2 y (mean)† Dense breasts: 100% Personal history: 9.8% Family history: NR
Hooley et al, 2012 (29) Fair	Test accuracy	United States	648 women	>15	Age: 52 y (mean)† Dense breasts: 100% Personal history: NR Family history: NR
Leong et al, 2012 (30) Fair	Test accuracy	Singapore	106 women	12–24	Age: 45.1 y (mean) Dense breasts: 100% Personal history: 5% Family history: 20.9%
Parris et al, 2013 (31) Fair	Cohort	United States	5519 women	NR	Age: 53.6 y (mean) Dense breasts: 89% Personal history: 6% Family history: 42%
Venturini et al, 2013 (32) Fair	Cohort	Italy	826 women	NR	Age: 100% <50 y† Dense breasts: 100% Personal history: NR Family history: 24%†
Weigert and Steenbergen, 2012 (33) Fair	Cohort	United States	8647 examinations 8647 women	NR	Age: 54.4 y (mean) Dense breasts: 100% Personal history: NR Family history: NR
Youk et al, 2011 (34) Fair	Test accuracy	South Korea	446 examinations	24	Age: 47.5 y (mean)† Dense breasts: 100% Personal history: 0% Family history: NR
Automated whole-breast ultrasonography					
Brem et al, 2015 (35) Fair	Cohort	United States	15 318 women	12	Age: 53.3 y (mean) Dense breasts: 100% Personal history: 3.6% Family history: 44.8%
Giuliano and Giuliano, 2013 (36) Fair	Cohort	United States	3418 women	12	Age: NR Dense breasts: 100% Personal history: 0% Family history: 0%

Continued on following page

Table 1–Continued

Study, Year (Reference) USPSTF Quality Rating	Design	Country	Examinations/ Women Analyzed	Follow-up Period, *mo*	Population Characteristics
Kelly et al, 2010 (37) Fair	Test accuracy	United States	6425 examinations 4419 women	12	Age: 53 y (mean)† Dense breasts: 68% Personal history: 10%† Family history: 30%
Magnetic resonance imaging					
Berg et al, 2012 (25) Good	Test accuracy	United States	334 examinations	12	Age: 56.8 y (mean) Dense breasts: 100% Personal history: NR Family history: NR
Kriege et al, 2006 (38) Good	Test accuracy	The Netherlands	1723 examinations	12	Age: 40 y (mean)† Dense breasts: 100% Personal history: NR Family history: 100%
Kuhl et al, 2014 (39) Good	Test accuracy	Germany	105 women	24	Age: 53.2 y (mean) Dense breasts: 100% Personal history: NR Family history: NR
Digital breast tomosynthesis					
Ciatto et al, 2013 (40) Fair	Cohort	Italy	1215 women	No follow-up except on biopsy results	Age: 58 y (median)† Dense breasts: 100% Personal history: NR Family history: NR
Haas et al, 2013 (41) Fair	Cohort	United States	4794 examinations	No follow-up except on biopsy results	Age: 33.3% <40 y† Dense breasts: 100% Personal history: 5.5%† Family history: NR
McCarthy et al, 2014 (42) Fair	Cohort	United States	8545 examinations	No follow-up except on biopsy results	Age: 70% >50 y Dense breasts: 100% Personal history: NR Family history: NR
Rose et al, 2013 (43) Fair	Cohort	United States	11 675 examinations	No follow-up except on biopsy results	Age: 54.2 y (mean)† Dense breasts: 100% Personal history: NR Family history: NR

NA = not applicable; NR = not reported; USPSTF = U.S. Preventive Services Task Force.
* Time between mammographic assessments.
† Data reflect the entire study population, not necessarily the subgroup with dense breasts.

tics of included studies. No studies addressed the effect of supplemental screening (compared with women without supplemental screening) on breast cancer morbidity or mortality.

Accuracy and Reliability of BI-RADS Density Determination

Absent a gold standard for breast density, studies could not evaluate the accuracy of BI-RADS density determinations. Five studies reported repeated assignment of categorical BI-RADS breast density classification by the same or different radiologists, altogether including more than 440 000 women, almost all with data from 2 sequential screening mammograms. To reflect current U.S. practice, we included only studies based on the BI-RADS density categories. The 3 largest studies were set in the United States. Two used data from the Breast Cancer Surveillance Consortium (20, 22), and the third presented findings from community radiologists conducting repeated readings of a large screening test set (24). Two other small studies (not discussed here) were based on mammographic screening programs in Spain (21) and Italy (23). All United States–based studies reflected community practice by use of clinical readings from community screening programs or test set readings by practicing community radiologists without additional training.

Overall, group prevalence of BI-RADS density ratings was similar across initial and subsequent examinations among community radiologists (**Appendix Table 2**, available at www.annals.org), but there was greater disagreement at the individual level. On subsequent screening examinations, approximately 1 in 5 women (23%) was placed in a different BI-RADS density category (a, b, c, d) by the same radiologist, while approximately 1 in 3 was categorized differently when a different radiologist read the subsequent examination result (**Table 2**). Considering clinical interpretations that combine categories ("dense" representing those with BI-RADS c or d and "nondense" representing BI-RADS a or b), 13% to 19% of women were reclassified into a different breast density category on their subsequent screening mammogram (**Table 2**).

These average estimates do not reflect greater extremes seen among outlier radiologists. Among 34

community radiologists reading sequential examination results in the same women (22), readers assigned the same BI-RADS density assessment on both mammograms 77% of the time, on average; however, individual readers' agreement between repeated ratings ranged from 62% to 87% (data not shown). In a study assessing repeat as well as cross-reader assignment of BI-RADS density categories by 19 radiologists in a test set of 341 examinations, radiologists assigned the same BI-RADS density assignment 82% of the time, on average, although individual readers varied from 66% to 95% (24).

In community settings, 19% to 22% of examinations initially classified as dense were subsequently reclassified as nondense, whereas 10% to 16% of initially nondense examinations were reclassified as dense (Table 2). In contrast, initial clinical readings for a test set showed a higher percentage reclassified from nondense to dense than vice versa. Across studies, the most commonly assigned breast density categories (b or c) were also those most likely to be reclassified on subsequent examination (Table 2), representing a clinical reclassification between nondense and dense. Radiologists tended to agree with their own previous assessments of density better than with those made by other readers, although there was substantial variability among pairs of readers due to outliers (more details in full report [16]). These results apply most to postmenopausal women or those aged 50 years and older because these women made up 71% to 100% of the study samples.

Test Performance Characteristics of Supplemental Screening Technologies in Women With Dense Breasts

Nine studies reported test performance characteristics for supplemental screening with HHUS, ABUS, and MRI among women with negative mammography results (Table 2 and Appendix Figures 3 and 4, available at www.annals.org). No studies reported test performance characteristics of DBT for women with dense breasts.

HHUS and ABUS

Two good-quality studies (from the United States [25] and Italy [26]) and 3 fair-quality studies (29, 30, 34) reported on HHUS, and 1 fair-quality study from the United States (37) reported on ABUS (Table 3). We found no studies reporting variation in performance of these modalities by patient age and other breast cancer risk factors among women with dense breasts. Both good-quality studies applied consistent reference standards to identify interval cancer and included more than 1000 women. The Italian study included women who self-referred to a charity-funded breast clinic and reported findings separately by breast density category. The U.S. study included only women with dense breasts, but many women also had additional major risk factors. The authors provided data for the subset of women without major risk factors. Additional details on all included studies are found in the full report (16).

Among women with dense breasts after recent negative results on screening mammography, the sensitivity of HHUS in the 2 good-quality studies for detecting all breast cancer (including ductal carcinoma in situ and invasive cancer) ranged from 0.80 (95% CI, 0.65 to 0.91) to 0.83 (CI, 0.59 to 0.96) (25, 26), and specificity ranged from 0.86 (CI, 0.85 to 0.88) to 0.95 (CI, 0.94 to 0.95) (25, 26). Sensitivity and specificity for invasive cancers were similar (25, 26). PPV in the good-quality studies ranged from 0.03 to 0.08; NPV was 0.99 (25, 26). A single fair-quality study found that ABUS had performance characteristics similar to those of HHUS among women with dense breasts and negative mammography results (37).

MRI

Three good-quality studies (25, 38, 39) reported test characteristics of supplemental MRI screening (Table 3). These studies included many women with elevated risk for breast cancer. In 2 studies, authors provided us with unpublished data for the subgroup of women with dense breasts, excluding women at very high risk because of *BRCA1/2* mutations, chest radiation, or personal histories of breast cancer (25, 39). In both, women had also recently had negative findings on screening with HHUS. The third study included stratified results based on risk factors (38).

Among these subgroups of lower-risk women with dense breasts, the sensitivity of MRI screening (after negative mammography results) for all breast cancer ranged across studies from 0.75 (CI, 0.35 to 0.97) to 1.00 (CI, 0.59 to 1.00) (25, 38, 39). Specificity also varied, ranging from 0.78 (CI, 0.73 to 0.83) (25) to 0.93 (CI, 0.87 to 0.97) (39). PPV ranged from 0.03 to 0.33 and NPVs were 0.99 to 1.00.

Cancer Detection and Recall Rates With Supplemental Screening

In general, supplemental screening after negative results on screening mammography consistently detected additional cases of breast cancer, most of which were invasive. Eighteen studies reported rates of additional cancer detected, and most also reported recall and biopsy rates associated with supplemental screening (Table 4 and Appendix Figure 5, available at www.annals.org). With the possible exception of DBT, supplemental testing led to many additional recalls and biopsies.

HHUS and ABUS

Seven studies reported HHUS cancer detection rates (27, 28, 31–33), and 3 studies reported on ABUS (35–37). The two good-quality studies of HHUS consistently estimated an all-cancer detection rate after negative mammography findings of 4.4 per 1000 examinations (CI, 2.5 to 7.2) (25, 26), with invasive cancer making up 93% (25) and 88% (26) of detected cancers. In the same women, mammography cancer detection rates were 4.7 per 1000 examinations in the U.S. study (25) and 2.8 per 1000 examinations in the Italian study

Table 2. Potential Misclassification of Breast Imaging Reporting and Data System Density Categorization by Density Categories

Study, Year (Reference) USPSTF Quality Rating	Repeat Examination Readers	Women Receiving a Different Breast Density Classification at Second Examination (4 categories), %	Women Receiving an Opposite Breast Density Classification at Second Examination (2 categories), %	Dense Examinations Reclassified as Nondense* (c or d to a or b), %
Harvey et al, 2013 (20) Good	Different community radiologists (*n* = 703)	32	18.7	22
Spayne et al, 2012 (22) Good	Same community radiologists (*n* = 34)	23	12.6	19
Gard et al, 2015 (24) Fair	Same community radiologists (*n* = 19)	29	16.9	10

USPSTF = U.S. Preventive Services Task Force.
* Categorized as "heterogeneously dense" or "extremely dense" at first examination and "almost entirely fat" or "scattered fibroglandular densities" at second examination.
† Categorized as "heterogeneously dense" at first examination and "almost entirely fat" or "scattered fibroglandular densities" at second examination.
‡ Categorized as "heterogeneously dense" at first examination and "scattered fibroglandular densities" at second examination.
§ Categorized as "almost entirely fat" or "scattered fibroglandular densities" at first examination and "heterogeneously dense" or "extremely dense" at second examination.
‖ Categorized as "scattered fibroglandular densities" at first examination and "heterogeneously dense" or "extremely dense" at second examination.
¶ Categorized as "scattered fibroglandular densities" at first examination and "heterogeneously dense" at second examination.

(26). Only the U.S. study reported the recall rate for supplemental HHUS: 14% (CI, 12.7% to 15.1%) (25).

Three fair-quality studies reported cancer detection rates for ABUS. Cancer detection rates after negative results on mammography ranged from 1.9 to 15.2 per 1000 examinations (36, 37). In comparison, the cancer detection rate from mammography alone in 1 of these studies was 4.3 per 1000 examinations (37). Recall rates varied between the studies from 2% (CI, 1.1% to 2.0%) to 14% (CI, 12.9% to 14.0%) (35, 36).

MRI

In 3 good-quality studies of MRI after negative mammography results, breast cancer detection rates varied from 3.5 (CI, 1.3 to 7.6) to 28.6 (CI, 5.9 to 81.2) per 1000 examinations (25, 38, 39), with small numbers of cancer cases detected (range, 2 to 7). In comparison, rates of mammography cancer detection in 2 of these studies for women with dense breasts were 4.1 and 7.0 per 1000 examinations (25, 38). Invasive breast cancer made up 67% and 86% of detected cancer, as reported by 2 studies (25, 39). Notably, women in these studies probably had higher breast cancer risk than the general population of women with dense breasts. A good-quality U.S. study evaluated supplemental HHUS and MRI among 334 women without *BRCA* mutations or previous breast cancer; after 3 screening rounds with negative mammography and HHUS results over 24 months, screening breast MRI identified 6 additional cases of invasive cancer (25).

Recall rates ranged from 9% (CI, 4.0% to 15.7%) to 23% (CI, 18.9% to 28.3%); the rate was highest in the study with 3 rounds of screening (25, 39). Biopsy rates were not reported separately for subgroups of women without increased risk. Because 2 of the studies reported on only 1 round of screening, the cumulative effect of recall for additional imaging and biopsy would likely increase with additional screening rounds.

DBT

Four fair-quality studies of DBT (3 in the United States [41–43] and 1 in Italy [40]) reported on screening populations of women with dense breasts. All U.S. studies were single-site, retrospective studies, and generally focused on outcomes before and after DBT introduction. In 1 study, breast cancer risk among women was described as above average (41); other studies did not report on risk factors (40, 42, 43). Three studies reported cancer detection rates with digital mammography alone ranging from 4.0 to 5.2 per 1000 examinations (40, 42, 43). With DBT, combined detection ranged from 5.4 (CI, 3.5 to 7.9) to 6.9 (CI, 4.8 to 9.6) per 1000 examinations (42, 43). A single study reported that 67% of cancer cases detected with combined DBT and mammography were invasive, the same proportion as with mammography alone (42). Recall rates with DBT in 3 retrospective U.S. studies ranged from 7% (CI, 6.2% to 7.7%) to 11% (CI, 10.0% to 11.7%), compared with 9% (CI, 8.4% to 11.0%) to 17% (CI, 15.0% to 18.2%) with digital mammography alone (41–43).

Harms of Breast Density Notification

Only 1 study, a good-quality Canadian randomized, controlled trial, examined the effects of notifying women with normal screening results that their mammograms showed dense breasts (44). Women randomly assigned to the intervention group (*n* = 285) received a report of their breast density with letters summarizing their mammography results and a pamphlet on breast cancer risk factors, including density. No supplemental screening was recommended. Women randomly assigned to the control group (*n* = 333) were notified of mammography results without information on breast density. At 4 weeks, more women in the intervention group had statistically significantly increased knowledge of breast density (25% in the intervention group vs. 8% in the control group) and were more likely to perceive themselves as having elevated

Table 2–Continued

Dense Examinations Reclassified as Nondense† (c to a or b), %	Dense Examinations Reclassified as Nondense‡ (c to b), %	Nondense Examinations Reclassified as Dense§ (a or b to c or d), %	Nondense Examinations Reclassified as Dense‖ (b to c or d), %	Nondense Examinations Reclassified as Dense¶ (b to c)
21	20	16	16	15
19	18	10	10	10
10	10	23	23	23

breast cancer risk. These differences did not persist at 6 months. Psychological distress, breast cancer worry, and preoccupation with breast cancer did not differ between groups.

Other Harms of Supplemental Screening

In studies of supplemental screening with HHUS and ABUS, more than 90% of positive test results were false-positive, and in MRI studies 66% to 97% of all positive test results were false-positives. Although no studies specifically addressed harms of supplemental screening in women with dense breasts, harms stemming from false-positive results are likely to be at least equivalent to those from mammography (45). We found no studies of whether focus on breast density distracts from assessment of other risk factors for breast cancer. Use of gadolinium contrast required for breast MRI has been associated with nephrogenic systemic fibrosis in patients with acute kidney injury or chronic kidney disease, but we found no reports of this adverse effect specifically related to breast MRI. The ACR recommends screening with serum creatinine before administration of gadolinium for those aged 60 years and older with hypertension, diabetes, or history of renal disease (46). Harms from DBT could come from additional breast radiation exposure (40–43, 47).

DISCUSSION

We examined the consistency of categorical BI-RADS breast density determinations in U.S. community practices because this is the system recommended by the ACR and written into most of the legislative mandates. According to large, community practice–based studies, BI-RADS density assessments at a population level were generally consistent across sequential examinations by the same or different readers, but there was important variability among readings for individual women. Approximately 80% of examinations received a b or c BI-RADS density assessment; these categories were also most likely to be reassessed differently, whether on a separate reading of the same examina-

Table 3. Test Performance Characteristics for Supplemental Hand-Held Ultrasonography, Automated Whole-Breast Ultrasonography, and Magnetic Resonance Imaging

Study, Year (Reference) USPSTF Quality Rating	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Hand-held ultrasonography				
Berg et al, 2012 (25) Good	0.83 (0.59–0.96)	0.86 (0.85–0.88)	0.03 (0.02–0.05)	1.00 (1.00–1.00)
Corsetti et al, 2011 (26) Good	0.80 (0.65–0.91)	0.95 (0.94–0.95)	0.07 (0.05–0.10)	1.00 (1.00–1.00)
Hooley et al, 2012 (29) Fair	1.00 (0.29–1.00)	0.77 (0.73–0.80)	0.02 (0.01–0.06)	1.00 (0.99–1.00)
Leong et al, 2012 (30) Fair	1.00 (0.16–1.00)	0.79 (0.70–0.86)	0.08 (0.02–0.26)	1.00 (0.96–1.00)
Youk et al, 2011 (34) Fair	1.00 (0.72–1.00)	0.72 (0.67–0.76)	0.08 (0.05–0.14)	1.00 (0.99–1.00)
Automated whole-breast ultrasonography				
Kelly et al, 2010 (37) Fair	0.68 (0.50–0.83)	0.92 (0.91–0.92)	0.04 (0.03–0.06)	1.00 (1.00–1.00)
Magnetic resonance imaging				
Berg et al, 2012 (25) Good	1.00 (0.59–1.00)	0.78 (0.73–0.83)	0.09 (0.04–0.17)	1.00 (0.99–1.00)
Kriege et al, 2006 (38) Good	0.75 (0.35–0.97)	0.89 (0.87–0.90)	0.03 (0.01–0.06)	1.00 (1.00–1.00)
Kuhl et al, 2014 (39) Good	1.00 (0.29–1.00)	0.94 (0.88–0.98)	0.33 (0.12–0.66)	1.00 (0.96–1.00)

NPV = negative predictive value; PPV = positive predictive value; USPSTF = U.S. Preventive Services Task Force.

Table 4. Breast Cancer Detection Outcomes for Supplemental HHUS, ABUS, MRI, and DBT

Study, Year (Reference) USPSTF Quality Rating	Cancer Cases Detected	Cancer Detection Rate (95% CI)	Recall Rate %, (95% CI)	Biopsy Rate %*, (95% CI)
HHUS				
Berg et al, 2012 (25) Good	15/3414 examinations	4.4 per 1000 examinations (2.5–7.2)	14 (12.7–15.1)	NR
Corsetti et al, 2011 (26) Good	32/7224 examinations	4.4 per 1000 examinations (3.0–6.2)	NR	6 (5.4–6.5)
Brancato et al, 2007 (27) Fair	2/5227 women	0.4 per 1000 women (0–1.4)	2 (1.7–2.5)	1 (0.9–1.6)
Girardi et al, 2013 (28) Fair	22/9960 women	2.2 per 1000 women (1.4–3.3)	NR	NR
Hooley et al, 2012 (29) Fair	3/648 women	4.6 per 1000 women (1.0–13.5)	24 (20.3–27.0)	7 (5.2–9.4)
Leong et al, 2012 (30) Fair	2/106 women	18.9 per 1000 women (1.7–50.3)	17†	13 (7.4–21.2)†
Parris et al 2013 (31) Fair	10/5519 women	1.8 per 1000 women (0.9–3.3)	NR	3 (2.8–3.8)
Venturini et al, 2013 (32) Fair	2/826 women	2.4 per 1000 women (0.3–8.7)	10 (7.5–11.7)	1 (0.6–2.2)
Weigert and Steenbergen, 2012 (33) Fair	25/8647 women	2.9 per 1000 women (1.9–4.3)	14 (13.1–14.5)	5 (4.4–5.3)
Youk et al, 2011 (34) Fair	11/446 examinations	24.7 per 1000 examinations (12.4–43.7)	14 (10.6–17.2)	11 (8.2–14.3)
ABUS				
Brem et al, 2015 (35) Fair	30/15 318 women	1.9 per 1000 examinations (1.3–2.8)	14 (12.9–14.0)	4 (3.4–4.0)
Giuliano and Giuliano, 2013 (36) Fair	DM + ABUS: 52/3418 women DM: 19/4076 women	DM + ABUS: 15.21 per 1000 women (11.4–19.9) DM: 4.7 per 1000 women (2.8–7.3)	2 (1.1–2.0)	NR
Kelly et al, 2010 (37) Fair	23/6425 examinations	3.6 per 1000 examinations (2.3–5.4)	9 (8.0–9.4)	NR
MRI				
Berg et al, 2012 (25) Good	7/334 examinations	21 per 1000 examinations (8.5–42.7)	23 (18.9–28.3)	NR
Kriege et al, 2006 (38) Good	6/1723 examinations	3.5 per 1000 examinations (1.3–7.6)	12 (10.0–13.1)	NR
Kuhl et al, 2014 (39) Good	3/105 women	28.6 per 1000 women (5.9–81.2)	9 (4.0–15.7)	NR
DBT				
Ciatto et al, 2013 (40) Fair	DBT + DM: 8/1215 examinations DM: 5/1215 examinations	DBT + DM: 6.6 per 1000 examinations (2.9–12.9) DM: 4.1 per 1000 examinations (1.3–9.6)	DBT + DM: 7 (5.2–8.1) DM: 7 (5.8–8.8)	NR
Haas et al, 2013 (41) Fair	NR	NR	DBT + DM: 10 (8.6–10.9) DM: 17 (15.0–18.2)	NR
McCarthy et al, 2014 (42) Fair	DBT + DM: 35/5056 examinations DM: 18/3489 examinations	DBT + DM: 6.9 per 1000 examinations (4.8–9.6) DM: 5.2 per 1000 examinations (3.1–8.1)	DBT + DM: 11 (10.0–11.7) DM: 13 (11.7–14.0)	NR
Rose et al, 2013 (43) Fair	DBT + DM: 25/4666 examinations DM: 28/7009 examinations	DBT + DM: 5.4 per 1000 examinations (3.5–7.9) DM: 4.0 per 1000 examinations (2.7–5.8)	DBT + DM: 7 (6.2–7.7) DM: 9 (8.4–11.0)	NR

ABUS = automated whole-breast ultrasonography; DBT = digital breast tomosynthesis; DM = digital mammography; HHUS = hand-held ultrasonography; MRI = magnetic resonance imaging; NR = not reported; USPSTF = U.S. Preventive Services Task Force.
* Biopsy rate includes needle aspiration, core needle, and open biopsies.
† Data are based on the 106 women with complete follow-up (out of 141 total).

tion or on a subsequent examination, and whether read by the same or a different reader. As a result, across studies a sizeable 13% to 19% of women (13–19) were reclassified from "nondense" to "dense" or vice versa. In these instances, mandated communications about elevated breast cancer risk or the need for additional clinical screenings could provide inconsistent information for the same woman in the span of 2 to 3 years.

Breast density findings can change because of multiple factors related to the woman being examined, the qualitative nature of the technique, and radiologist variability in interpretation of the examinations. The studies we examined tried to control for within-woman biological factors, suggesting that most of the variation in breast density assessment reflects within- and between-radiologist variability in density interpretation and the limitations of the current BI-RADS approach. Concerns about BI-RADS breast density determinations are a major impetus for research examining other methods for assigning breast density, including automated volumet-

ric estimates, ultrasonographic assessments, and other computer-assisted methods. Although variability is reduced by use of double readings, which is widely practiced in Europe (40), this approach is impractical in the United States because of workforce requirements. The introduction of standards and quality measures related to breast density categorization could help to minimize potential harms associated with variable breast density categorizations.

When combined with mandated direct-to-consumer communications, variability in breast density assignments may lead to unintended consequences. Reclassification from one overall category to another (for example, "dense" to "not-dense" or vice versa) may undermine a woman's confidence in the screening process and leave her uncertain about her risk for breast cancer, whereas the opposite reclassification may alarm women unnecessarily or prompt supplemental screening tests of uncertain value. The ACR has publicly expressed similar cautions about benefits, possible harms, and unintended consequences for the communication of breast density assessments to women (48).

Few studies evaluated test performance of supplemental screening tests for women with dense breasts. In the studies identified, the sensitivity of supplemental MRI screening after negative screening mammography results appeared generally higher than that seen with HHUS screening. However, although we examined subsets of women without specific risk factors, we suspect that, in general, these women were at higher risk. Studies of MRI were small and variable in their sensitivity estimates. No study directly compared sensitivity of supplemental screening modalities among women with dense breasts. Specificity of supplemental screening modalities was similar, and PPV was low. We identified only one study of ABUS and no studies of DBT test performance in women with dense breasts. No studies examined the effects of age or other breast cancer risk factors on supplemental test performance characteristics in women with dense breasts. No studies reported on breast cancer morbidity and mortality outcomes.

Evidence on harms of supplemental screening was also sparse. Added to digital mammography, DBT more than doubles the radiation exposure from each screening examination (49–51). New estimates of cancer induced by radiation from breast imaging have recently been reported (47). Technology that allows reconstruction of the 2-dimensional breast images can reduce radiation exposure but is not widely disseminated (49). We found no reports of adverse effects from use of gadolinium contrast for breast MRI, but a tracking mechanism for this potentially severe, albeit rare, adverse effect should be considered. Potential harms resulting from overdiagnosis of breast cancer through supplemental screening can be identified only through rigorous prospective studies with long-term follow-up.

Our review was limited to studies published in English; studies published in other languages may have met inclusion criteria, although applicability to U.S. practice could be limited. For applicability and feasibility concerns, we focused only on BI-RADS breast density assessment. Studies did not examine the underlying reasons for variability in BI-RADS assessment within or between radiologists, nor did they evaluate any interventions to reduce the variability. The number, quality, and rigor of studies of diagnostic test characteristics and clinical outcomes were limited. Most studies lacked a complete reference standard, sufficient follow-up, or a clear description of follow-up, so diagnostic test performance characteristics could not be evaluated. Recall was often not clearly defined. No studies compared interval breast cancer rates, stage at diagnosis, or breast cancer mortality among two groups of women with dense breasts undergoing screening mammography with or without supplemental testing. No studies addressed the important potential risks of overdiagnosis and the associated harms of unnecessary treatment. Many studies included mixtures of women at increased breast cancer risk due to risk factors other than breast density, limiting the generalizability to the general screening population of women with dense breasts. Literature on ABUS and DBT for women with dense breasts was limited, as was literature on the harms of breast density notification. Only 1 comparative study of cohorts with and without supplemental screening adjusted for differences between cohorts (42).

In conclusion, good-quality studies with U.S. radiologists show important reclassification between dense and nondense breasts in women undergoing sequential screening examinations. Reclassification of breast density may introduce confusion or reduce confidence among women. Moving from a "dense" to a "nondense" breast categorization may result in different mandated communications in states with breast density notification, as well as fluctuation in clinical recommendations for supplemental screening.

Limited evidence suggests that more breast cancer cases will be detected by supplemental HHUS and MRI screening of women with dense breasts, and most detected breast cancer cases will be invasive. Studies have not evaluated whether diagnosis of additional breast cancer by supplemental screening leads to improved clinical outcomes or what proportion of the cancer diagnosed represents overdiagnosis. Supplemental testing of women with dense breasts with HHUS or MRI is associated with increased recall rates for diagnostic investigation among women without breast cancer. Use of DBT may be associated with lower recall rates, but studies are few and retrospective. To define meaningful clinical outcomes of supplemental screening of women with dense breasts, well-designed, long-term, prospective, comparative studies of supplemental screening are needed.

From the Center for Healthcare Policy and Research, University of California, Davis, Sacramento, California; Group Health Research Institute, Group Health Cooperative, Seattle, Washington; Kaiser Permanente Research Affiliates, NW, Portland, Oregon; and Columbia University, New York, New York.

Disclaimer: This review was conducted by the Kaiser Permanente Research Affiliates Evidence-based Practice Center with

the University of California Davis Center for Healthcare Policy and Research under contract to AHRQ. AHRQ staff provided oversight for the project and assisted in the external review of the companion draft evidence synthesis. The analytic framework, review questions, and methods for locating and qualifying evidence were posted on the USPSTF website for public comment before the review began; final versions reflect public input. The authors of this report are responsible for its content, including any clinical treatment recommendations. No statement in this article should be construed as an official position of AHRQ or the U.S. Department of Health and Human Services.

Acknowledgment: The authors thank the following for their contributions to this project: AHRQ staff; the USPSTF; Joann Elmore, MD, MPH, Elizabeth Rafferty, MD, Jeffrey Tice, MD, Edward Sickles, MD, Barnett Kramer, MD, MH, Gretchen Gierach, PhD, and Gwendolyn Bryant-Smith, MD, who provided expert and federal partner review of the report; Wendie Berg, MD, PhD, and Christiane Kuhl, MD, for providing unpublished subgroup data; and Bruce Abbott, MLS, and Guibo Xing, PhD, at the University of California, Davis.

Financial Support: By AHRQ (contract HHSA-290-2012-00015I).

Disclosures: Dr. Melnikow reports a contract with the Agency for Healthcare Research and Quality during the conduct of the study. Dr. Miglioretti reports grants from the Agency for Healthcare Research and Quality and the National Cancer Institute during the conduct of the study. Ms. Weyrich reports grants from the Agency for Healthcare Research and Quality during the conduct of the study. Ms. Thompson reports grants from the Agency for Healthcare Research and Quality during the conduct of the study. Authors not named here have disclosed no conflicts of interest. Disclosures can also be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M15-1789.

Requests for Single Reprints: Reprints are available from the AHRQ Web site (www.ahrq.gov).

Current author addresses and author contributions are available at www.annals.org.

References

1. **Sprague BL, Gangnon RE, Burt V, Trentham-Dietz A, Hampton JM, Wellman RD, et al.** Prevalence of mammographically dense breasts in the United States. J Natl Cancer Inst. 2014;106. [PMID: 25217577] doi:10.1093/jnci/dju255

2. **Carmona-Sánchez E, Cuadros López JL, Cuadros Celorrio ÁM, Pérez-Roncero G, González Ramírez AR, Fernández Alonso AM.** Assessment of mammographic density in postmenopausal women during long term hormone replacement therapy. Gynecol Endocrinol. 2013;29:1067-70. [PMID: 24004297] doi:10.3109/09513590.2013.831831

3. **del Carmen MG, Hughes KS, Halpern E, Rafferty E, Kopans D, Parisky YR, et al.** Racial differences in mammographic breast density. Cancer. 2003;98:590-6. [PMID: 12879477]

4. **Centers for Disease Control and Prevention.** Breast cancer rates by race and ethnicity. 2014. Accessed at www.cdc.gov/cancer/breast/statistics/race.htm on 3 March 2015.

5. **Gierach GL, Ichikawa L, Kerlikowske K, Brinton LA, Farhat GN, Vacek PM, et al.** Relationship between mammographic density and breast cancer death in the Breast Cancer Surveillance Consortium. J Natl Cancer Inst. 2012;104:1218-27. [PMID: 22911616] doi:10.1093/jnci/djs327

6. **Lee CH, Dershaw DD, Kopans D, Evans P, Monsees B, Monticciolo D, et al.** Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. J Am Coll Radiol. 2010;7:18-27. [PMID: 20129267] doi:10.1016/j.jacr.2009.09.022

7. **National Comprehensive Cancer Network.** Breast cancer screening and diagnosis. 2014. Accessed at www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf on 10 October 2014.

8. **American Cancer Society.** American Cancer Society recommendations for early breast cancer detection in women without breast symptoms. 2014. Accessed at www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs on 7 October 2014.

9. **Committee on Gynecologic Practice.** ACOG Committee Opinion no. 593: Management of women with dense breasts diagnosed by mammography. Obstet Gynecol. 2014;123:910-1. [PMID: 24785636] doi:10.1097/01.AOG.0000445584.44898.7d

10. Are You Dense? D.E.N.S.E. State Efforts. 2015. Accessed at www.areyoudenseadvocacy.org/dense on 29 June 2015.

11. Illinois General Assembly. (215 ILCS 5/356g) Sec. 356g. Mammograms; mastectomies. Accessed at www.ilga.gov/legislation/ilcs/documents/021500050k356g.htm on 10 October 2014.

12. 118th General Assembly of the State of Indiana. Senate Enrolled Act No. 414. 2013. Accessed at www.in.gov/legislative/bills/2013/PDF/SE/SE0414.1.pdf on 29 June 2015.

13. The State of Connecticut. Public Act No. 09-41. An Act Requiring Communication of Mammographic Breast Density Information to Patients. 2009. Accessed at www.cga.ct.gov/2009/ACT/PA/2009PA-00041-R00SB-00458-PA.htm on 10 October 2015.

14. New Jersey 215th Legislature. An Act Concerning Mammograms, amending P.L.1991, c.279 and P.L.2004, c.86, and supplementing Title 26 of the Revised Statutes and P.L.2003, c.193. 2014. Accessed at www.njleg.state.nj.us/2012/Bills/PL13/196_.PDF on 10 October 2014.

15. Breast Density and Mammography Reporting Act of 2015. 114th Congress of the United States ed. 2015. Accessed at https://www.congress.gov/bill/114th-congress/senate-bill/370 on 10 October 2014.

16. **Melnikow J, Fenton J, Whitlock E, Miglioretti DL, Weyrich M, Thompson J, et al.** Adjunctive screening for breast cancer in women with dense breasts: a systematic review for the U.S. Preventive Services Task Force. AHRQ publication no. 14-05201-EF-2. Rockville, MD: Agency for Healthcare Research and Quality. [Forthcoming]

17. **U.S. Preventive Services Task Force.** U.S. Preventive Services Task Force Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality; 2013.

18. **National Institute for Health and Clinical Excellence.** The Guidelines Manual. London: National Institute for Health and Clinical Excellence; 2006.

19. **Lucas NP, Macaskill P, Irwig L, Bogduk N.** The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). J Clin Epidemiol. 2010;63:854-61. [PMID: 20056381] doi:10.1016/j.jclinepi.2009.10.002

20. **Harvey JA, Gard CC, Miglioretti DL, Yankaskas BC, Kerlikowske K, Buist DS, et al; Breast Cancer Surveillance Consortium.** Reported mammographic density: film-screen versus digital acquisition. Radiology. 2013;266:752-8. [PMID: 23249570] doi:10.1148/radiol.12120221

21. **Redondo A, Comas M, Macià F, Ferrer F, Murta-Nascimento C, Maristany MT, et al.** Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. Br J Radiol. 2012;85:1465-70. [PMID: 22993385] doi:10.1259/bjr/21256379

22. **Spayne MC, Gard CC, Skelly J, Miglioretti DL, Vacek PM, Geller BM.** Reproducibility of BI-RADS breast density measures among community radiologists: a prospective cohort study. Breast J. 2012;18:326-33. [PMID: 22607064] doi:10.1111/j.1524-4741.2012.01250.x

23. **Bernardi G, Cavallaro G, Indinnimeo M, Fiore A, Basso L, D'Ermo G, et al.** Usefulness of ultrasounds in the management of breast phyllodes tumors. G Chir. 2012;33:81-5. [PMID: 22525552]

24. **Gard CC, Aiello Bowles EJ, Miglioretti DL, Taplin SH, Rutter CM.** Misclassification of Breast Imaging Reporting and Data System (BI-RADS) mammographic density and implications for breast density reporting legislation. Breast J. 2015;21:481-9. [PMID: 26133090] doi:10.1111/tbj.12443

25. **Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al; ACRIN 6666 Investigators.** Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. JAMA. 2012;307:1394-404. [PMID: 22474203] doi:10.1001/jama.2012.388

26. **Corsetti V, Houssami N, Ghirardi M, Ferrari A, Speziani M, Bellarosa S, 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. Eur J Cancer. 2011;47:1021-6. [PMID: 21211962] doi:10.1016/j.ejca.2010.12.002

27. **Brancato B, Bonardi R, Catarzi S, Iacconi C, Risso G, Taschini R, et al.** Negligible advantages and excess costs of routine addition of breast ultrasonography to mammography in dense breasts. Tumori. 2007;93:562-6. [PMID: 18338490]

28. **Girardi V, Tonegutti M, Ciatto S, Bonetti F.** Breast ultrasound in 22,131 asymptomatic women with negative mammography. Breast. 2013;22:806-9. [PMID: 23558244] doi:10.1016/j.breast.2013.02.010

29. **Hooley RJ, Greenberg KL, Stackhouse RM, Geisel JL, Butler RS, Philpotts LE.** Screening US in patients with mammographically dense breasts: initial experience with Connecticut Public Act 09-41. Radiology. 2012;265:59-69. [PMID: 22723501]

30. **Leong LC, Gogna A, Pant R, Ng FC, Sim LS.** Supplementary breast ultrasound screening in Asian women with negative but dense mammograms-a pilot study. Ann Acad Med Singapore. 2012;41:432-9. [PMID: 23138139]

31. **Parris T, Wakefield D, Frimmer H.** Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. Breast J. 2013;19:64-70. [PMID: 23240937] doi:10.1111/tbj.12053

32. **Venturini E, Losio C, Panizza P, Rodighiero MG, Fedele I, Tacchini S, et al.** Tailored breast cancer screening program with microdose mammography, US, and MR Imaging: short-term results of a pilot study in 40-49-year-old women. Radiology. 2013;268:347-55. [PMID: 23579052] doi:10.1148/radiol.13122278

33. **Weigert J, Steenbergen S.** The Connecticut experiment: the role of ultrasound in the screening of women with dense breasts. Breast J. 2012;18:517-22. [PMID: 23009208] doi:10.1111/tbj.12003

34. **Youk JH, Kim EK, Kim MJ, Kwak JY, Son EJ.** Performance of hand-held whole-breast ultrasound based on BI-RADS in women with mammographically negative dense breast. Eur Radiol. 2011;21:667-75. [PMID: 20853108] doi:10.1007/s00330-010-1955-8

35. **Brem RF, Tabár L, Duffy SW, Inciardi MF, Guingrich JA, Hashimoto BE, et al.** Assessing improvement in detection of breast cancer with three-dimensional automated breast US in women with dense breast tissue: the SomoInsight Study. Radiology. 2015;274:663-73. [PMID: 25329763] doi:10.1148/radiol.14132832

36. **Giuliano V, Giuliano C.** Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. Clin Imaging. 2013;37:480-6. [PMID: 23116728] doi:10.1016/j.clinimag.2012.09.018

37. **Kelly KM, Dean J, Comulada WS, Lee SJ.** Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur Radiol. 2010;20:734-42. [PMID: 19727744] doi:10.1007/s00330-009-1588-y

38. **Kriege M, Brekelmans CT, Obdeijn IM, Boetes C, Zonderland HM, Muller SH, et al.** Factors affecting sensitivity and specificity of screening mammography and MRI in women with an inherited risk for breast cancer. Breast Cancer Res Treat. 2006;100:109-19. [PMID: 16791481]

39. **Kuhl CK, Schrading S, Strobel K, Schild HH, Hilgers RD, Bieling HB.** Abbreviated breast magnetic resonance imaging (MRI): first postcontrast subtracted images and maximum-intensity projection-a novel approach to breast cancer screening with MRI. J Clin Oncol. 2014;32:2304-10. [PMID: 24958821] doi:10.1200/JCO.2013.52.5386

40. **Ciatto S, Houssami N, Bernardi D, Caumo F, Pellegrini M, Brunelli S, et al.** Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol. 2013;14:583-9. [PMID: 23623721] doi:10.1016/S1470-2045(13)70134-7

41. **Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE.** Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. Radiology. 2013;269:694-700. [PMID: 23901124] doi:10.1148/radiol.13130307

42. **McCarthy AM, Kontos D, Synnestvedt M, Tan KS, Heitjan DF, Schnall M, et al.** Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. J Natl Cancer Inst. 2014;106. [PMID: 25313245] doi:10.1093/jnci/dju316

43. **Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr.** Implementation of breast tomosynthesis in a routine screening practice: an observational study. AJR Am J Roentgenol. 2013;200:1401-8. [PMID: 23701081] doi:10.2214/AJR.12.9672

44. **Bottorff JL, Ratner PA, Johnson JL, Hislop TG, Buxton JA, Zeisser C, et al.** Women's responses to information on mammographic breast density. Can J Nurs Res. 2007;39:38-57. [PMID: 17450704]

45. **Nelson HD, Cantor A, Humphrey L, Fu R, Pappas M, Daeges M, et al.** Screening for breast cancer: a systematic review to update the 2009 U.S. Preventive Services Task Force recommendation. AHRQ publication no. 14-05201-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; 2015.

46. **Besheli LD, Aran S, Shaqdan K, Kay J, Abujudeh H.** Current status of nephrogenic systemic fibrosis. Clin Radiol. 2014;69:661-8.

47. **Miglioretti DL, Lange J, van Ravesteyn N, van den Broek JJ, Lee CI, Melnikow J, et al.** Radiation-induced breast cancer and breast cancer death from mammography screening. AHRQ publication no. 14-05201-EF-5. Rockville, MD: Agency for Healthcare Research and Quality. [Forthcoming]

48. **American College of Radiology.** ACR statement on reporting breast density in mammography reports and patient summaries. 24 April 2012. Accessed at www.acr.org/About-Us/Media-Center/Position-Statements/Position-Statements-Folder/Statement-on-Reporting-Breast-Density-in-Mammography-Reports-and-Patient-Summaries on 24 April 2013.

49. **Gur D, Zuley ML, Anello MI, Rathfon GY, Chough DM, Ganott MA, et al.** Dose reduction in digital breast tomosynthesis (DBT) screening using synthetically reconstructed projection images: an observer performance study. Acad Radiol. 2012;19:166-71. [PMID: 22098941] doi:10.1016/j.acra.2011.10.003

50. **Feng SS, Sechopoulos I.** Clinical digital breast tomosynthesis system: dosimetric characterization. Radiology. 2012;263:35-42. [PMID: 22332070] doi:10.1148/radiol.11111789

51. **Olgar T, Kahn T, Gosch D.** Average glandular dose in digital mammography and breast tomosynthesis. Rofo. 2012;184:911-8. [PMID: 22711250]

Current Author Addresses: Drs. Melnikow and Fenton and Ms. Weyrich: Center for Healthcare Policy and Research, University of California, Davis, 2103 Stockton Boulevard, Sacramento, CA 95817.
Dr. Whitlock and Ms. Thompson: Kaiser Permanente Center for Health Research, 3800 North Interstate Avenue, Portland, OR 97227.
Dr. Miglioretti: Department of Public Health Sciences, University of California Davis School of Medicine, One Shields Avenue, Med Sci 1C, Room 145, Davis, CA 95616.
Mr. Shah: Columbia University, 6380 Lerner Hall, 2920 Broadway, New York, NY 10027.

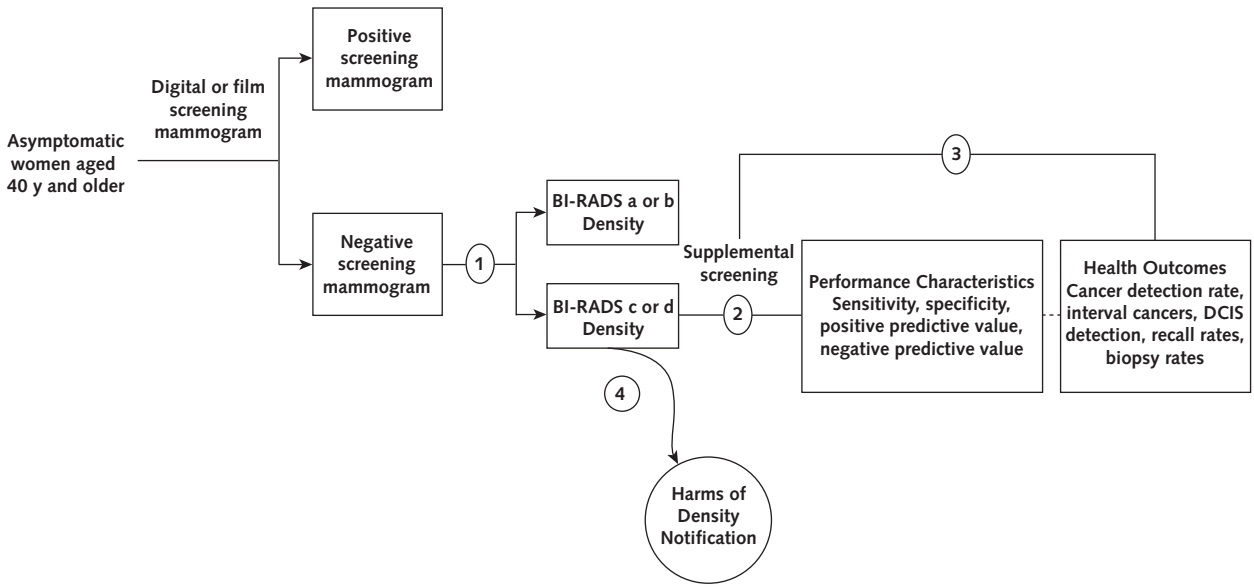
Author Contributions: Conception and design: J. Melnikow, J.J. Fenton, E.P. Whitlock, D.L. Miglioretti, K. Shah.
Analysis and interpretation of the data: J. Melnikow, J.J. Fenton, E.P. Whitlock, D.L. Miglioretti, M.S. Weyrich, J.H. Thompson, K. Shah.
Drafting of the article: J. Melnikow, J.J. Fenton, M.S. Weyrich, J.H. Thompson, K. Shah.
Critical revision of the article for important intellectual content: J. Melnikow, J.J. Fenton, E.P. Whitlock, D.L. Miglioretti, M.S. Weyrich, J.H. Thompson, K. Shah.
Final approval of the article: J. Melnikow, J.J. Fenton, E.P. Whitlock, D.L. Miglioretti, M.S. Weyrich, J.H. Thompson, K. Shah.
Statistical expertise: D.L. Miglioretti.
Obtaining of funding: J. Melnikow, E.P. Whitlock.
Administrative, technical, or logistic support: J. Melnikow, D.L. Miglioretti, M.S. Weyrich, J.H. Thompson, K. Shah.
Collection and assembly of data: J. Melnikow, J.J. Fenton, E.P. Whitlock, D.L. Miglioretti, M.S. Weyrich, J.H. Thompson, K. Shah.

Appendix Table 1. Breast Density Legislation in the United States

Status of Legislation*	Legislative Details	States
Pending Legislation	Drafting legislation mandating breast density notification	Florida, Maine, Illinois, Colorado, Vermont, Mississippi
	Introduced legislation mandating breast density notification†	Washington, Iowa, Indiana, Kentucky, South Carolina, Georgia
Enacted Legislation	Mandates patient notification about breast density	California, Arizona, Oregon, Nevada, Massachusetts, Minnesota, Texas, Alabama, Missouri, Tennessee, North Carolina, Virginia, Maryland, New Jersey, Pennsylvania, New York, Connecticut, Rhode Island, Hawaii, Michigan, Ohio, Louisiana, Delaware, North Dakota
	Requires specific language for patient notification	California, Arizona, Texas, Alabama, Missouri, Tennessee, North Carolina, Virginia, Maryland, New Jersey, Pennsylvania, New York, Connecticut, Rhode Island, Hawaii, Ohio, Michigan, Louisiana, Massachusetts
	Requires that all mammography reports provide information about breast density and the patient's current breast density level	Nevada, North Carolina, Maryland, Pennsylvania, Connecticut, Louisiana
	Requires that insurers cover appropriate medical examinations and tests for women with dense breasts	Illinois, Connecticut, New Jersey, Indiana

Source: (10).
* As of September 2015.
† During the 2015 legislative season.

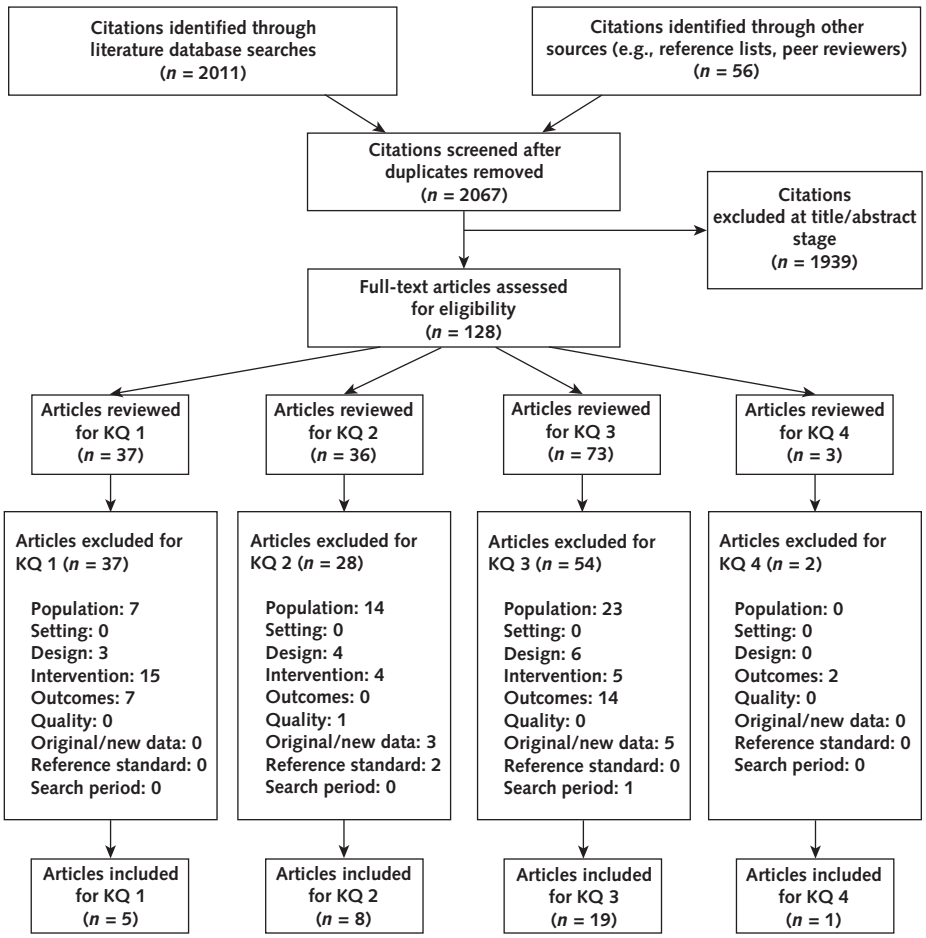
Appendix Figure 1. Analytic framework.



KQ 1: What are the accuracy and reproducibility of BI-RADS determination of breast density?
KQ 2: What are the test performance characteristics of newer technologies for breast cancer screening when used as supplemental tests after a negative screening mammogram in women found to have dense breasts, and how do these performance characteristics differ by age and risk factors?
KQ 3: When performed after a negative screening mammogram in women found to have dense breasts, what is the effectiveness of supplemental screening with breast ultrasonography, MRI, or breast tomosynthesis on proximate clinical outcomes, including cancer detection rates, DCIS detection rates, stage at diagnosis, recall rates, biopsy rates, and interval cancer rates?
KQ 4: What are the harms associated with being identified as having dense breasts, including psychological and quality-of-life impacts and harms associated with supplemental screening evaluation, including evaluation of false-positive results?

BI-RADS = Breast Imaging Reporting and Data System; DCIS = ductal carcinoma in-situ; KQ = key question; MRI = magnetic resonance imaging.

Appendix Figure 2. Summary of evidence search and selection.



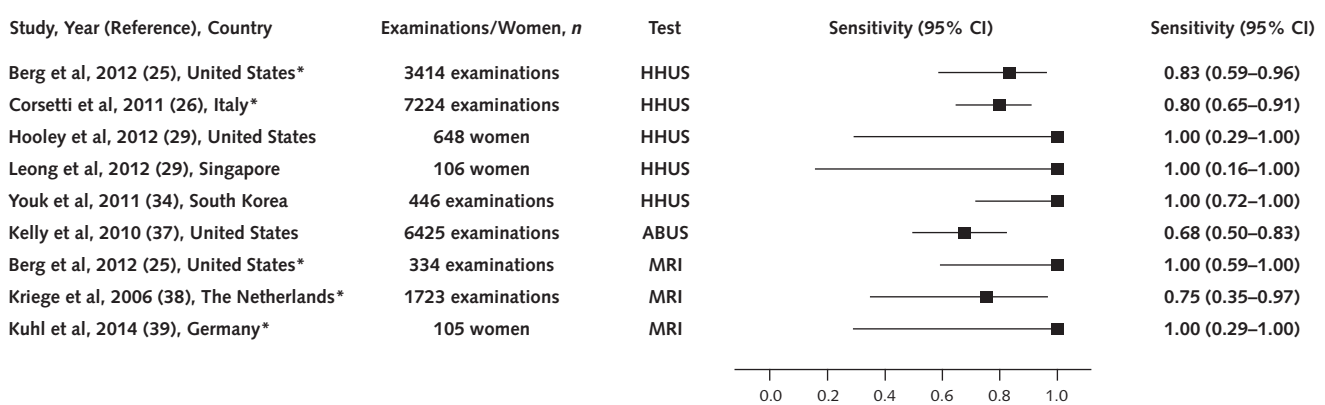
KQ = key question.

Appendix Table 2. Consistency of Breast Imaging Reporting and Data System Density Categories and Population Categorization

Study Time Between Assessments	Study Sample Inclusion Criteria	Assessments	a* (%)	b† (%)	c‡ (%)	d§ (%)
Harvey et al, 2013 (20) <36 mo	Women aged 40 years and older, with no history of breast cancer or reported use of hormone therapy at the time of exam or during the previous year, and had two or more screening mammographic examinations less than 36 months apart between January 1, 2000, and December 31, 2009 (n = 435,751)	Exam 1 Exam 2	9.4 10.2	45.2 45.1	37.9 37.2	7.5 7.2
Spayne et al, 2012 (22) 3-24 mo	Women who were postmenopausal‖, with no history of breast cancer or reported use of hormone therapy, and had two or more film-screen screening or bilateral diagnostic mammograms including BI-RADS breast density assessments between January 1, 1996 and December 31, 2006 (n = 11,755)	Exam 1 Exam 2	9.8 9.2	61.0 60.2	26.6 28.1	2.5 2.5
Gard et al, 2015 (24) 6 mo	Women contributing examinations to the test set had a screening mammogram interpreted in the health care system between 1996 and 1998 and were enrolled in the system for at least 2 years following screening; the test set was designed to include about twice as many examinations of women with cancer as without, and roughly equal numbers of non-dense and dense examinations based on clinical interpretation (n = 341)	Reading 1¶ Reading 2	6.1 4.5	44.3 39.2	38.3 47.0	11.4 9.3

BI-RADS=Breast Imaging Reporting and Data System.
* Breast density category a = almost entirely fat.
† Breast density category b = scattered fibroglandular densities.
‡ Breast density category c = heterogeneously dense.
§ Breast density category d = extremely dense.
‖ Aged 55 years or older or reported having experienced natural menopause, having had both ovaries removed, or having more than 365 days elapse since their last menstrual period.
¶ Clinical interpretation of exams prior to inclusion in test set.
# Based on BI-RADS classification by the majority of readers for each exam in the test set.

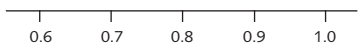
Appendix Figure 3. Sensitivity of supplemental screening with hand-held ultrasonography, automated whole-breast ultrasonography, and magnetic resonance imaging in detecting breast cancer.



These estimates include ductal carcinoma in situ and invasive cancers. ABUS = automated whole-breast ultrasonography; HHUS = hand-held ultrasonography; MRI = magnetic resonance imaging.
* Good-quality study.

Appendix Figure 4. Specificity of supplemental screening with hand-held ultrasonography, automated whole-breast ultrasonography, and magnetic resonance imaging in detecting breast cancer.

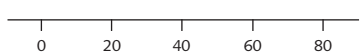
Study, Year (Reference), Country	Examinations/Women, *n*	Test	Specificity (95% CI)	Specificity (95% CI)
Berg et al, 2012 (25), United States*	3414 examinations	HHUS		0.86 (0.85–0.88)
Corsetti et al, 2011 (26), Italy*	7224 examinations	HHUS		0.95 (0.94–0.95)
Hooley et al, 2012 (29), United States	648 women	HHUS		0.77 (0.73–0.80)
Leong et al, 2012 (29), Singapore	106 women	HHUS		0.79 (0.70–0.86)
Youk et al, 2011 (34), South Korea	446 examinations	HHUS		0.72 (0.67–0.76)
Kelly et al, 2010 (37), United States	6425 examinations	ABUS		0.92 (0.91–0.92)
Berg et al, 2012 (25), United States*	334 examinations	MRI		0.78 (0.73–0.83)
Kriege et al, 2006 (38), The Netherlands*	1723 examinations	MRI		0.89 (0.87–0.90)
Kuhl et al, 2014 (39), Germany*	105 women	MRI		0.94 (0.88–0.98)



These estimates include ductal carcinoma in situ and invasive cancer. ABUS = automated whole-breast ultrasonography; HHUS = hand-held ultrasonography; MRI = magnetic resonance imaging.
* Good-quality study.

Appendix Figure 5. Breast cancer detection rates of supplemental screening with hand-held ultrasonography, automated whole-breast ultrasonography, magnetic resonance imaging, and digital breast tomosynthesis.

Study, Year (Reference), Country	Examinations/Women, *n*	Test	Rate per 1000 (95% CI)	Rate per 1000 (95% CI)
Berg et al, 2012 (25), United States*	3414 examinations	HHUS		4.39 (2.46–7.24)
Corsetti et al, 2011 (26), Italy*	7224 examinations	HHUS		4.43 (3.03–6.25)
Brancato et al, 2007 (27), Italy	5227 examinations	HHUS		0.38 (0.05–1.38)
Girardi et al, 2013 (28), Italy	9960 women	HHUS		2.21 (1.39–3.34)
Hooley et al, 2012 (29), United States	648 women	HHUS		4.63 (0.96–13.47)
Leong et al, 2012 (30), Singapore	106 women	HHUS		14.18 (1.72–50.30)
Parris et al, 2013 (31), United States	5519 women	HHUS		1.81 (0.87–3.33)
Venturini et al, 2013 (32), Italy	826 women	HHUS		2.42 (0.29–8.72)
Weigert and Steenberger, 2012 (33), United States	8647 women	HHUS		2.89 (1.87–4.27)
Youk et al, 2011 (34), South Korea	446 examinations	HHUS		24.66 (12.38–43.70)
Brem et al, 2015 (35), United States	15318 women	ABUS		1.96 (1.32–2.79)
Giuliano and Giuliano, 2013 (36), United States	3418 women	ABUS		15.21 (11.40–19.90)
Kelly et al, 2010 (37), United States	6425 examinations	ABUS		3.58 (2.27–5.37)
Berg et al, 2012 (25), United States*	334 examinations	MRI		20.96 (8.47–42.70)
Kriege et al, 2006 (38), The Netherlands*	1723 examinations	MRI		3.48 (1.28–7.56)
Kuhl et al, 2014 (39), Germany*	105 women	MRI		28.57 (5.93–81.23)
Ciatto et al, 2013 (40), Italy	1215 examinations	DBT		6.58 (2.85–12.93)
McCathy et al, 2014 (42), United States	5056 examinations	DBT		6.92 (4.83–9.61)
Rose et al, 2013 (43), United States	4666 examinations	DBT		5.36 (3.47–7.90)



These estimates include ductal carcinoma in situ and invasive cancer. ABUS = automated whole-breast ultrasonography; DBT = digital breast tomosynthesis; HHUS = hand-held ultrasonography; MRI = magnetic resonance imaging.
* Good-quality study.