

Evidence Synthesis

Number 126

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

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
www.ahrq.gov

Contract No. HHS-290-2012-00015-I, Task Order No. 5

Prepared by:

Center for Healthcare Policy and Research
University of California, Davis
Sacramento, CA

Kaiser Permanente Research Affiliates Evidence-Based Practice Center
Kaiser Permanente Center for Health Research
Portland, OR

Investigators:

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
Kunal Shah

AHRQ Publication No. 14-05201-EF-3
January 2016

This report is based on research conducted by the University of California, Center for Healthcare Policy and Research and the Kaiser Permanente Research Affiliates Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HHS-2012-00015-I, Task Order No. 5). The findings and conclusions in this document are those of the authors, who are responsible for its contents, and do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information (i.e., in the context of available resources and circumstances presented by individual patients).

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors acknowledge the following individuals for their contributions to this project: Bruce Abbott, MLS, for creating and conducting the literature searches; Wendy Berg, MD, PhD, Zheng Zhang, PhD, and Christiane K. Kuhl, MD, for providing subgroup data; Guibo Xing, PhD, for calculation of confidence intervals and creation of the graphs; Karen Lindfors, MD, MPH, for consultation on clinical and technical aspects of supplemental screening; Joann Elmore, MD, MPH, Elizabeth Rafferty, MD, Jeffrey Tice, MD, Edward Sickles, MD, and Barnett Kramer, MD, MPH, who provided expert review of the report; Gwendolyn Bryant-Smith, MD, and Gretchen L. Gierach, PhD, who provided federal partner review of this report; Jennifer Croswell, MD, MPH, at AHRQ; and current and former members of the U.S. Preventive Services Task Force who contributed to topic deliberations.

Suggested Citation

Melnikow JM, Fenton JJ, Whitlock EP, Miglioretti DM, Weyrich MS, Thompson JS, Shah K. Supplemental Screening for Breast Cancer in Women With Dense Breasts: A Systematic Review for the U.S. Preventive Services Task Force. Evidence Synthesis No. 126. AHRQ Publication No. 14-05201-EF-3. Rockville, MD: Agency for Healthcare Research and Quality; 2016.

Structured Abstract

Background: Screening mammography has lower sensitivity and specificity for women with increased breast density, who also have a higher risk of breast cancer.

Purpose: To systematically review the evidence on the accuracy and reproducibility of the Breast Imaging-Reporting and Data System (BI-RADS) breast density assessment scores, as well as the evidence on test performance and clinical outcomes of supplemental screening with hand-held ultrasound (HHUS), automated whole breast ultrasound (ABUS), magnetic resonance imaging (MRI), and digital breast tomosynthesis (DBT) for women with dense breasts and negative screening mammography.

Data Sources: We searched MEDLINE, PubMed, Embase, and the Cochrane library from January 2000 through July 2015. We reviewed the reference lists of included studies and relevant systematic reviews to identify relevant articles that were published before the timeframe or not identified in our literature searches. We also searched the grey literature for relevant reports and reviewed their references, and identified articles based on suggestions from experts. We searched Clinicaltrials.gov to identify relevant ongoing trials.

Study Selection: Two reviewers independently reviewed the titles and abstracts of all identified articles to determine if studies met the inclusion and exclusion criteria. All studies were required to report the study population and results for women with BI-RADS breast density c/d or equivalent. Two reviewers then independently evaluated the potential relevant full-text articles against a priori inclusion and exclusion criteria. Disagreements in the abstract and/or full-text review were resolved through consensus discussion.

Data Extraction: A single reviewer independently abstracted study characteristics and results into tables. A second reviewer independently reviewed each study and checked tables for accuracy. Subgroups of women with dense breasts were abstracted separately when reported or if data were provided by study authors.

Data Analysis: Evidence for all key questions was qualitatively synthesized. Sensitivity, specificity, positive and negative predictive value, cancer detection rates, recall rates, and biopsy rates were calculated for individual study subgroups of women with dense breasts. 95% confidence intervals were calculated by the exact method for each study estimate of sensitivity, specificity, cancer detection rates, and biopsy rates.

Results: There is no recognized gold standard for breast density determination, so no studies were identified that evaluated the accuracy of the BI-RADS breast density categories in screening mammography. Five studies reported statistical measures for the reproducibility of categorical BI-RADS breast density classification among women predominantly or exclusively receiving screening mammograms. Best estimates from U.S. data suggest about one in five women would be categorized into a different BI-RADS density category (a, b, c, d) by the same radiologist at the next screening exam, while one in three would be categorized differently if the next screening exam were read by a different radiologist. Major re-categorization (i.e., from “dense” categories [c or d] to “non-dense” categories [a or b], or vice versa) at the next screening

examination occurred in 12.6 to 18.7 percent of women.

For test performance characteristics of supplemental screening of women with dense breasts and negative screening mammography, two good-quality and three fair-quality studies reported on HHUS, one fair-quality study reported on ABUS, and three good-quality studies reported on MRI. We identified no studies of DBT performance among women with dense breasts and negative screening mammography. In the good-quality HHUS studies, for all breast cancer (defined as including DCIS and invasive breast cancer) the sensitivity ranged from 80.0 to 83.0 percent, specificity ranged from 86.4 to 94.5 percent and positive predictive value (PPV) ranged from 3.2 to 7.5 percent. For ABUS, the sensitivity was 67.6 percent, specificity was 91.6 percent, and PPV was 4.1 percent. In the three MRI studies, which were smaller and included high-risk women, the sensitivity ranged from 75.0 to 100.0 percent, specificity ranged from 78.1 to 88.7 percent and PPV ranged from 3.0 to 33.3 percent.

No studies were identified that examined the impact of supplemental screening on breast cancer recurrence rates or mortality for women with dense breasts. We identified observational studies that reported breast cancer detection rates, recall rates, and biopsy rates: ten studies of HHUS (two good-quality), three fair-quality studies of ABUS, three good-quality studies of MRI, and four fair-quality studies of DBT. Most studies compared screening outcomes in the same cohort pre- and post-supplemental testing. One study of HHUS, two of ABUS and three studies of DBT compared clinical outcomes of two groups of women undergoing mammography, with and without supplemental testing. Supplemental testing consistently found additional breast cancers not identified by mammography, but increased false positive results, with the possible exception of DBT. The two good-quality studies of HHUS had consistent estimates of the incremental (additional after mammography) cancer detection rate: 4.4 per 1,000 exams. In the good-quality U.S. study, recall rates for additional imaging and/or biopsies were 139 per 1,000 exams; in the good-quality Italian study, the biopsy rate was 59 per 1,000 exams. In two fair-quality studies of ABUS, the cancer detection rates were 4.6 and 1.9 per 1,000 exams and recall was 87 and 150 per 1,000 exams. For MRI, incremental cancer detection rates ranged from 3.5 to 28.6 per 1,000 exams. Recall rates for additional diagnostic testing ranged from 115 to 235 per 1,000 exams. For DBT, cancer detection rates rose from 4.0 to 4.1 breast cancers per 1,000 exams with digital mammography alone to 5.4 to 6.6 breast cancers per 1,000 exams with added DBT. Recall rates declined with the addition of DBT in all studies: from 91 to 69 per 1,000 exams; from 72 to 66 per 1,000 exams, from 128 to 108 per 1,000 exams, and from 166 to 97 per 1,000 exams. Across all modalities, invasive cancers (rather than ductal carcinomas in-situ) comprised 89 to 93 percent of cancers detected by HHUS, 74 to 93 percent of cancers detected by ABUS, 67 to 86 percent of cancers detected by MRI, and 68 to 92 percent of those detected by DBT.

We identified one RCT comparing potential harms of notification of breast density to a control group. No differences in psychological outcomes or intention for clinical breast exam were detected at 6 months. We found no studies on potential harms of receiving different breast density classification on sequential examinations. Harms of supplemental screening with ultrasound or MRI of women with dense breasts include higher recall and biopsy rates when compared with digital mammography alone. Harms of breast MRI include risk of nephrogenic systemic fibrosis for women with advanced chronic kidney disease. DBT use in conjunction with digital mammography more than doubles the radiation exposure of each combined screening

exam.

Limitations: Studies of BI-RADS reproducibility may reflect somewhat older community practice. No studies examined long term outcomes of supplemental screening for women with dense breasts. Many studies of test performance and proximate clinical outcomes were of fair-quality and most were conducted in cohorts of women with risk factors in addition to dense breasts. Six observational studies compared cohorts with and without supplemental screening, but only one employed statistical techniques to adjust for differences in baseline risk between groups.

Conclusions: Reproducibility of BI-RADS density determinations in U.S. community practice does not appear to be ideal. Mammograms from 12.6 to 18.7 percent of women were reclassified into a different overall combined category (i.e., from “non-dense” to “dense” or vice versa) at their next screening exam when read by the same or a different radiologist, which may introduce confusion or reduce confidence among women receiving mandated breast density notifications. This would affect certainty of any recommendation for supplemental screening of women identified as having dense breasts. Studies identifying more accurate and reproducible methods of identifying women with dense breasts are needed. There were no published studies of important longer-term clinical outcomes of supplemental screening. In general, supplemental screening of women with dense breasts will lead to the identification of more breast cancers (mostly invasive), but may be associated with higher recall rates and additional biopsies. Whether cancers identified by supplemental screening have better outcomes and how many of them represent cancers that would not otherwise become clinically apparent (overdiagnosis) cannot be determined from the studies published to date. Rigorous comparative studies of supplemental screening for women with dense breasts including clinical outcomes beyond breast cancer diagnosis are needed for all modalities.

Table of Contents

Chapter 1. Introduction	1
Purpose.....	1
Background.....	1
Condition Definition	1
Prevalence and Burden of Condition	2
Risk Factors for Dense Breasts	3
Current Clinical Practice in the United States	3
Screening Strategies	3
Clinical Guidelines	4
Breast Density Legislation	4
Previous USPSTF Recommendation	5
Chapter 2. Methods	6
Scope and Purpose	6
Key Questions and Analytic Framework	6
Data Sources and Searches	6
Study Selection	7
Quality Assessment and Data Extraction.....	7
Data Synthesis and Analysis.....	9
Expert Review and Public Comment.....	9
USPSTF Involvement.....	10
Chapter 3. Results.....	11
Literature Search.....	11
Key Question 1. What Is the Accuracy and Reproducibility of BI-RADS Determination of Breast Density?	11
Summary of Results	11
Study Details.....	12
Key Question 2. What Are the Test Performance Characteristics of Newer Technologies for Breast Cancer Screening When Used as Supplemental Tests After a Negative Screening Mammography Exam in Women Found to Have Dense Breasts? How Do These Performance Characteristics Differ by Age and Risk Factors?.....	15
Summary of Results	15
Study Details.....	17
Key Question 3. When Performed After a Negative Screening Mammogram in Women Found to Have Dense Breasts, What Is the Effectiveness of Supplemental Screening With Breast Ultrasound, MRI, or DBT on Proximate Clinical Outcomes, Including Cancer Detection Rates, DCIS Detection Rates, Stage at Diagnosis, Recall Rates, Biopsy Rates, and Interval Cancer Rates?.....	20
Summary of Results	20
Study Details.....	22
Key Question 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?	28
Summary of Results.....	28
Study Details.....	29

Chapter 4. Discussion	31
Summary of Evidence.....	31
BI-RADS Breast Density Determinations	31
Supplemental Test Performance and Outcomes for Women With Dense Breasts and a Negative Mammogram	32
Harms of Breast Density Notification and Supplemental Testing.....	33
Limitations of the Review.....	33
Limitations Due to Our Approach	33
Limitations Due to the Evidence Base.....	33
Future Research Needs	34
Conclusion	34
References.....	35

Figures

- Figure 1. Analytic Framework
- Figure 2. Sensitivity of Supplemental HHUS, ABUS, and MRI in Detecting All Breast Cancer
- Figure 3. Specificity of Supplemental HHUS, ABUS, and MRI in Detecting All Breast Cancer
- Figure 4. Sensitivity of Supplemental HHUS and MRI in Detecting Invasive Breast Cancer
- Figure 5. Specificity of Supplemental HHUS and MRI in Detecting Invasive Breast Cancer
- Figure 6. Breast Cancer Detection Rates of Supplemental HHUS, ABUS, MRI, and DBT
- Figure 7. Invasive Breast Cancer Detection Rates of Supplemental HHUS, ABUS, MRI, and
DBT
- Figure 8. Biopsy Rates of Supplemental HHUS and ABUS

Tables

- Table 1. BI-RADS Breast Density Descriptions
- Table 2. Breast Density Legislation in the United States
- Table 3. U.S. Studies of BI-RADS Density Categories and Consistency of Population
Categorization
- Table 4. U.S. Studies of Potential Misclassification of BI-RADS Density Categorization for
Women and by Density Categories
- Table 5. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts:
Study and Population Characteristics
- Table 6. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Study
and Population Characteristics
- Table 7. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Study
and Population Characteristics
- Table 8. Supplemental DBT Screening for Breast Cancer in Women With Dense Breasts: Study
and Population Characteristics
- Table 9. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts:
Screening Test Characteristics
- Table 10. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts:
Screening Test Characteristics
- Table 11. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts:
Screening Test Characteristics

Table 12. Supplemental DBT Screening for Breast Cancer in Women With Dense Breasts: Screening Test Characteristics

Table 13. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Test Performance Characteristics

Table 14. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Test Performance Characteristics

Table 15. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Test Performance Characteristics

Table 16. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Table 17. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Table 18. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Table 19. Supplemental DBT Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Table 20. Cancer Detection, Recall, and Biopsy Rates of Good-Quality Supplemental Screening Studies

Table 21. Harms Associated With Breast Density Notification: Study and Population Characteristics

Table 22. Harms Associated With Breast Density Notification: Knowledge, Perceived Risk, and Breast Cancer Worry

Table 23. Summary of the Evidence

Chapter 1. Introduction

Purpose

The U.S. Preventive Services Task Force (USPSTF) will use this report, along with the corresponding report on breast cancer screening in the general population, to update its 2009 recommendation on breast cancer screening.

Background

Condition Definition

The appearance of the breast on mammography varies due to differences in breast tissue composition. Dense breasts are identified based on mammographic appearance; compared to non-dense breasts, dense breasts have more fibroglandular tissue (resulting in white areas on mammography) and less fatty tissue (which appears as dark areas on mammography). Wolfe first described an association between mammographic density and breast cancer risk in 1976.¹

Several systems of breast density classification have been developed since Wolfe described a categorization of the degree of breast density, defining four groups based on qualitative assessment of visual appearance. To date, computer assisted assessment of breast density has been used primarily as a research tool.² In the United States, breast density is generally defined and classified by visual assessment according to the American College of Radiology's (ACR) Breast Imaging-Reporting and Data System (BI-RADS) four-category scale.

Instituted by the American College of Radiology in 1992, BI-RADS provides a lexicon of breast imaging descriptors, a structure for reporting assessment categories and management recommendations, and a framework for data collection and auditing.³ The original BI-RADS document included only a summary of breast density and a statement describing general breast tissue type to address emerging evidence that increased breast density was associated with decreased mammographic sensitivity.³ The first edition of BI-RADS, published in 1992 and updated in 1998, introduced the four category classification of breast density (ranging from “almost entirely fat” [category 1] to “extremely dense” [category 4]) to improve and standardize communication of predicted mammographic performance and breast cancer risk.^{3,4} In 2003, the fourth edition of BI-RADS added quartile ranges of percentage dense tissue to each of the four density categories (i.e., 0-25% glandular tissue; approximately 26-50% glandular; approximately 51-75% glandular; >75% glandular tissue). The purpose was to align more closely with research on percent density assessments and to move towards distributing population breast density assessments more evenly across categories.⁵ Today, the fifth and current edition of the BI-RADS Atlas (2013) reverts to qualitative breast density reporting categories and emphasizes the importance of dense tissue masking some non-calcified cancers as being clinically important (**Table 1**).⁵ Breast density is reported as an overall assessment and not separately for each breast.

The 2013 BI-RADS density terminology classifies levels of breast density and is used to standardize reporting of visual assessment of mammograms.^{5,6} The categories employ the letters “a” to “d”, in place of numbers (to avoid confusion with the BI-RADS assessment scores classifying abnormalities) and categorize breast density in the following way:⁵

- a) The breasts are almost entirely fatty
- b) There are scattered areas of fibroglandular density
- c) The breasts are heterogeneously dense, which may obscure small masses
- d) The breasts are extremely dense, which lowers the sensitivity of mammography

Although the descriptions and naming of the categories have changed slightly between versions, the number of categories has been constant and overall distinctions between them have not shifted substantially across versions. All of the research included in this systematic review began before the transition to BI-RADS 2013 density terminology and hence, dense breasts were generally categorized as BI-RADS density 3 and 4; in the report, these are converted to c and d to match the current system.

Prevalence and Burden of Condition

Based on an analysis of over 3.8 million screening mammograms collected by the Breast Cancer Screening Consortium (BCSC)⁵ and using both pre- and post-2003 BI-RADS density guidelines, radiologist-assigned distribution of density was designated as:

- 10 percent of patients receiving mammograms are classified as having fatty breasts
- 40 percent as having scattered areas of density
- 40 percent as having heterogeneous density
- 10 percent as having extremely dense breasts

A subsequent analysis of over 1.5 million mammograms in the United States between 2007 and 2010 collected by the BCSC found that 43.3 percent of women aged 40 to 74 years had heterogeneous or extremely dense breasts (BI-RADS categories c or d). In the U.S. population, an estimated 27.6 million women fall into these categories.⁷

High breast density has two distinct effects. Abnormal areas on mammograms are more difficult to see in areas with dense breast tissue, reducing the sensitivity and specificity of mammography. Abnormalities on mammograms are more difficult to see in areas with dense breast tissue: sensitivity decreased from 81.2 percent to 63.3 percent and specificity decreased from 93.5 percent to 88.7 percent in women aged 40 to 49 years; sensitivity decreased from 90.0 percent to 57.1 percent and specificity from 95.0 percent to 93.8 percent in women aged 70 to 74 years.⁸ In addition, increased breast density is a significant predictor of increased breast cancer risk,⁹ with some studies estimating a four- to six-fold increase in lifetime breast cancer risk among women with extremely dense breasts relative to women with entirely fatty breasts.¹⁰⁻¹³ Comparisons of relative risk between the lowest and highest density categories may be misleading as these categories include only approximately 20 percent of women; more clinically relevant comparisons of relative risk between the two ‘dense’ categories and average breast density estimate increases in lifetime breast cancer risk of only 1.2- and 2.1-fold.¹⁴ A recent meta-

analysis of 13 case-control studies where density was measured on a computerized estimate of absolute dense area found a summary adjusted odds ratio for breast cancer risk of 1.37 (95% CI, 1.29 to 1.47) per standard deviation increase in absolute dense area on mammography in premenopausal women. The odds ratio was adjusted for age, body mass index (BMI), and parity. The summary adjusted odds ratio for the post-menopausal women was similar (OR 1.38; 95% CI, 1.31 to 1.44).¹⁵ Increased breast density is not, however, associated with higher breast cancer mortality among women with dense breasts diagnosed with breast cancer.¹⁶

Risk Factors for Dense Breasts

Breast density declines with age and body mass index (BMI), and increases with HRT use. A BCSC study of 764,507 women aged 40 years and older undergoing 1,518,599 mammograms evaluated the relationship between age, BMI and breast density and projected estimates for the population of women potentially eligible for breast cancer screening in the United States. Among the estimated 27.6 million women with dense breasts (BI-RADS c or d) aged 40 to 74 years, nearly 45 percent were between the ages of 40 and 49 years. In the BCSC population, the relative risk of breast cancer for women with dense breasts (BI-RADS c or d) compared to those without dense breasts (BI-RADS a or b) was 1.83 (unpublished data). Twenty five percent of obese women (BMI \geq 30) had heterogeneous or extremely dense breasts, compared to 40 percent of overweight women (BMI 25.0 to 29.9) and 58 percent of women with a normal BMI (BMI 18.5 to 24.9).⁷ HRT use is associated with higher breast density.¹⁷ In a separate large study, women on HRT with low/normal BMI and extremely dense breasts were shown to have the highest risk of breast cancer (OR 1.49; 95% CI, 1.21 to 1.83) relative to women not on HRT.¹⁸

Few differences in breast density by race/ethnicity have emerged, but evidence is limited. Two studies in the United States have found higher average breast density in Asian women. A study using BCSC data with over one million women found the proportion of Asian women with extremely dense breasts was highest across all age groups.¹⁹ A study of over 15,000 women attending breast cancer screening at Massachusetts General Hospital found that breast density on average was higher in Asian women, after adjusting for age and BMI.²⁰ In contrast, a study from the UK of 5,277 mammograms taken from 645 women found that after adjustment for age and BMI, South Asian women had 3.8 percent (95% CI, 1.1 to 6.3) lower average breast density.²¹ Of note, Asian women in the United States have lower than average incidence of breast cancer.²²

Current Clinical Practice in the United States

Screening Strategies

Because higher mammographic breast density reduces the sensitivity and specificity of breast cancer screening,⁸ and increased breast density is also associated with a higher risk of breast cancer,⁹⁻¹³ the use of supplemental breast cancer screening with additional screening modalities has been proposed as a method for better identification of breast cancers.¹⁴ The four most common supplemental screening modalities to mammography examined in the literature and implemented in clinical practice are hand-held breast ultrasound, automated whole breast ultrasound, breast MRI, and digital breast tomosynthesis.

Hand-held breast ultrasound (HHUS) is the systematic application of an ultrasound probe over each breast by a radiologist or breast imaging technologist. It is widely available and does not involve exposure to radiation. Automated whole breast ultrasonography (ABUS) involves placement of an ultrasound transducer on the breast by a technologist. The automated probe then moves over the breast in a standard fashion until the breast is scanned, storing 3,000 to 5,000 images for later review by a radiologist. Several systems are FDA approved and their use is becoming more common, but the technology is not widely available.

Breast magnetic resonance imaging (MRI) creates detailed MR images of the breasts. After the injection of IV gadolinium, breast MRI requires scanning of the breasts while lying still inside an MRI scanner. A typical breast MRI lasts 30 to 60 minutes and generates several hundred images.²³ Breast MRI has been recommended for annual supplemental screening of women with high (greater than 20%) estimated lifetime risk of breast cancer by the American Cancer Society, the American College of Radiology, and the National Comprehensive Cancer Network.²⁴⁻²⁶

Digital breast tomosynthesis (DBT) uses a computer algorithm to reconstruct multiple low-dose digital images of the breast into thin “slices” spanning the entire breast. These images can be displayed individually or in cine mode. Use of DBT is expanding as concurrent supplemental screening with mammography. In 2013, the FDA approved the use of synthetic 2-D images to take place of the standard 2-D digital mammogram; although it is currently not known how often synthetic views are used. The use of DBT in addition to standard mammography more than doubles the total radiation exposure.²⁷⁻³⁰ Use of newer synthetic 2-D image reconstruction to replace digital mammography eliminates the additional radiation of a digital mammogram, so that the radiation dose is due only the DBT exam.³¹ Several systems for synthetic 2-D images have been FDA approved, based on studies showing similar performance of synthetic 2-D images with full-field digital mammography^{32, 33} but these systems are not yet in widespread use.

Clinical Guidelines

Currently, there are no clinical guidelines explicitly recommending the use of supplemental breast cancer screening in women with dense breasts. The American College of Radiology (ACR)/Society of Breast Imaging²⁶ recommends “considering” supplemental HHUS for women with dense breasts, but notes the concerns about HHUS performance and resource limitations given the large number of women who would be potential candidates for screening. The National Comprehensive Cancer Network (NCCN)³⁴ cites insufficient evidence to recommend its use in women with dense breasts and no other risk factors. The American Cancer Society (ACS)³⁵ and NCCN both cite insufficient evidence to recommend for or against MRI screening as an adjunct to mammography in women with dense breasts. NCCN also cites insufficient evidence to recommend the use of DBT. The American College of Obstetricians and Gynecologists (ACOG) does not recommend the use of any supplemental screening tests in asymptomatic women with dense breasts and no other risk factors.³⁶

Breast Density Legislation

Legislation in many states requires that providers notify patients regarding breast density, and in

some cases, requires insurance coverage of subsequent supplemental screening. Legislation is currently pending in additional states and at the federal level. As of July 2015, twenty-four states, encompassing more than half the population of the United States, have enacted legislation requiring breast density information be included on each patient's mammography results report; an additional six states have introduced similar legislation and another six states are drafting legislation (**Table 2**).³⁷ Most legislation defines dense breasts as those being categorized as "heterogeneously" or "extremely dense" according to BI-RADS, while other state legislation employ more vague definitions, such as "consistent with current medical evidence"³⁸ or lack a definition altogether. Of the twenty-four states mandating notification, six require that all mammography reports include information about breast density and the patient's BI-RADS density classification, regardless of whether the patient has dense breasts.³⁷ Of the twenty-four states mandating notification, nineteen states' legislation specifies language that the healthcare provider must use, and usually includes information such as confirmation of dense breast tissue; BI-RADS density classification and explanation; impact of density on mammography-detected breast cancer; suggestions for further supplemental screening. Three states – Illinois,³⁹ Connecticut,⁴⁰ and New Jersey⁴¹ – have enacted legislation mandating that all insurers cover medical examinations and tests for women with dense breasts and Indiana⁴² mandates such coverage for women covered by state employee health insurance. At the federal level, legislation has been introduced in 2014⁴³ and 2015⁴⁴ that would require breast density notification in all mammography reports.

Previous USPSTF Recommendation

The 2009 USPSTF recommendation statement on screening for breast cancer focused on the general population, and not specifically on screening women with dense breasts and a negative mammogram.⁴⁵ The recommendation acknowledged that digital mammography had increased sensitivity (compared with film mammography) for women with dense breasts. It called for future randomized trials comparing the effectiveness of digital versus film mammography for women with dense breasts. Such trials have not been conducted and digital mammography is presently used for over 95 percent of mammograms in the United States.⁴⁶

Chapter 2. Methods

Scope and Purpose

This systematic review focused on the accuracy and reliability of breast density categorization, and on supplemental imaging tests for screening for women with dense breasts and negative mammography. The review complements the systematic review on breast cancer screening in the general population. This review summarized the current evidence regarding the reliability of density classification, the test performance characteristics of supplemental tests after a negative screening mammogram, the evidence on clinical outcomes of supplemental screening in women found to have dense breasts, and the harms associated with identification of dense breasts and supplemental screening in this population.

Key Questions and Analytic Framework

Using the USPSTF's methods⁴⁷ (detailed in **Appendix A**), we developed an analytic framework (**Figure 1**) and four key questions (KQs):

1. What is the accuracy and reproducibility of BI-RADS determination of breast density?
2. What are the test performance characteristics of newer technologies for breast cancer screening when used as supplemental tests after a negative screening mammogram exam in women found to have dense breasts and how do these performance characteristics differ by age and risk factors?
3. When performed after a negative screening mammogram in women found to have dense breasts, what is the effectiveness of supplemental screening with breast ultrasound, 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?
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?

Data Sources and Searches

The literature search for this systematic review included searches of MEDLINE, PubMed, Embase, and the Cochrane library from January 2000 through July 2015. We worked with a medical librarian to develop our search strategy (**Appendix A**). The literature search results were managed using version X7.1 of EndNote® (Thomson Reuters, New York, NY).

To ensure the comprehensiveness of our retrieval strategy, we reviewed the reference lists of included studies and relevant systematic reviews to identify relevant articles that were published before the timeframe or not identified in our literature searches. We also supplemented our database searches with suggestions from experts, and searched the grey literature for relevant reports related to supplemental screening of women with dense breasts. We reviewed their

references of relevant articles and reports, and searched Clinicaltrials.gov to identify relevant ongoing trials (**Appendix B**).

Study Selection

Two reviewers independently reviewed the titles and abstracts of identified articles to determine if studies met the inclusion and exclusion criteria for design, population, intervention, and outcomes (**Appendix A Table 1**). Two reviewers then independently evaluated the full-text article(s) of potentially included studies against the complete inclusion and exclusion criteria. Disagreements in the abstract and/or full-text review were resolved by discussion. Excluded full-text articles and reasons for exclusion are listed in **Appendix C**.

We developed an *a priori* set of criteria for inclusion and exclusion criteria of studies based on our understanding of the literature (**Appendix A Table 1**). For KQ 1 examining the intra-rater concordance and inter-rater reliability of BI-RADS density determination, we considered RCTs, prospective and retrospective cohort studies, case-control studies, cross-sectional studies, and test sets involving multiple blinded readings of digital or film mammography by at least three readers. For KQ 2 examining the test performance characteristics of supplemental screening for dense breasts with MRI, hand-held ultrasound, whole breast ultrasound, and DBT, we considered RCTs, prospective and retrospective cohort studies, and diagnostic accuracy studies. We required that the studies reported on a screening practice with two or more radiologists and that a reference standard of a minimum of 12 months clinical followup be applied to all participants with a negative screen. For KQ 3, examining the proximate clinical outcomes of supplemental breast cancer screening, we considered RCTs, prospective and retrospective cohort studies (including cohort studies reporting on cancer detection rates, stage distribution of detected cancers, recall rates and biopsy rates) and meta-analyses. We required studies to report outcomes from two or more radiologists. Study populations and outcomes reported had to be stratified by breast density in the screened group if women with all breast densities were included, so that rates for women with dense breasts could be calculated. KQ 4 examined the harms of being identified as having dense breasts; we considered RCTs, prospective and retrospective cohort studies, case-control studies, cross-sectional studies, meta-analyses, and modeling studies.

Quality Assessment and Data Extraction

Two reviewers independently assessed the methodological quality of each study using predefined criteria developed by the USPSTF⁴⁸ and supplemented with the National Institute for Health and Clinical Excellence methodology checklists⁴⁹ and the QAREL tool⁵⁰ for assessing diagnostic reliability. Disagreements in quality were resolved by discussion. Each study was given a final quality rating of good, fair, or poor.

Good-quality diagnostic reliability studies used a representative sample of subjects and raters, had blinded assessment of the reference standard (where applicable) and also blinded raters to non-clinical cues and to others ratings, used a varied examination order, an appropriate time interval between repeated measures, appropriate approaches to application and interpretation of

the test, and used appropriate statistical measures of agreement. Diagnostic reliability studies were downgraded to fair if they were unable to meet the majority of good-quality criteria. Poor-quality observational studies had multiple threats to internal validity and were excluded from this review (**Appendix C**).

Good-quality diagnostic accuracy studies applied credible reference standards interpreted independently of the screening test to both positive and negative test results (typically tissue sampling for positive results and minimum 12 month clinical followup for negative results, had low rates of loss to followup, and included at least 100 patients undergoing breast cancer screening. Studies were rated as poor if they did not apply a reference standard to the entire study sample including those with negative tests, had biased ascertainment of the reference standard, a very small sample size or limited the sample to very high risk patients, or had other threats to internal validity. Poor studies were excluded from the review (**Appendix C**).

Good-quality RCTs had adequate randomization procedures and allocation concealment, blinded outcome assessment, reliable outcome measures, similar groups at baseline (i.e., small differences between groups in baseline demographics and characteristics), low attrition ($\geq 90\%$ of participants had followup data, with < 10 percentage-point differences in loss to followup between groups), and used conservative data substitution methods if missing data were inferred. Poor-quality studies were excluded from the review (**Appendix C**).

Good-quality observational studies had adequate, unbiased ascertainment of exposed and unexposed groups. These studies addressed a population without the outcome of interest at the beginning of the study, and they had reliable outcome measures, blinded assessment, low attrition, adjustment for potential confounders, and no other important threats to internal validity. Observational studies were downgraded to fair if they were unable to meet the majority of good-quality criteria. Poor-quality observational studies had multiple threats to internal validity and were excluded from this review. Observational studies reviewed for this report often reported on diagnostic test outcomes, so criteria for diagnostic accuracy studies were applied when relevant (**Appendix C**).

One reviewer extracted data from all included studies rated as fair- or good-quality into a standard evidence table. A second reviewer checked the data for accuracy. For all KQs, elements abstracted included population characteristics (e.g., baseline demographics, breast density, family or personal history of breast cancer), study design (e.g., inclusion/exclusion criteria, followup, screening rounds), and screening test characteristics (e.g., reference standard, number of readers, radiological experience). For KQ 1, we abstracted details of inter- and intra-rater reliability and variability. For KQ 2, we abstracted test performance characteristics, including sensitivity, specificity, negative predictive value, and positive predictive value. For KQ 3, we abstracted proximate health outcomes, including breast cancer detection rates, invasive breast cancer detection rates, recall rates, and biopsy rates. Recall rates were abstracted as including recall for both true and false positive results. Recall rates were not always specified as recall for further imaging separately from biopsy, but biopsy rates were recorded separately when available. For KQ 4, we considered studies on adverse events associated with screening tests, and psychological harms of breast density notification to women.

Data Synthesis and Analysis

We created summary evidence tables to synthesize data separately for each KQ. The tables were the basis of our qualitative synthesis. In the context of study quality, we examined the range of results and looked for possible associations between study results and population or modality characteristics. Sensitivity, specificity, positive and negative predictive value, cancer detection rates, recall rates, and biopsy rates were calculated for subgroups of women with dense breasts when required. 95% confidence intervals were calculated with the exact method^{51,52} for each study estimate of sensitivity, specificity, cancer detection rates, and biopsy rates. In good-quality studies, when reported outcomes data were not stratified for women with dense breasts, we requested from authors study data stratified by breast density to enable inclusion of these studies in our review.

Expert Review and Public Comment

A draft version of this report was peer-reviewed by five invited content experts as well as federal partners from the National Cancer Institute and the U.S. Department of Veterans Affairs (VA). Comments received during this process were presented to the USPSTF during its deliberation of the evidence and subsequently addressed, as summarized below, in this version of the report. Additionally, a full draft report was posted for public comment on the USPSTF's website from April 20 to May 18, 2015. All comments were reviewed and considered. Most comments related to the report addressed 6 major areas:

- 1. Clarification on development of the BIRADS density system and other methods of assessing breast density**
Text on the development of the BIRADS density classification system was revised and other approaches to breast density assessment were mentioned in the *Introduction*.
- 2. Clarification about the relationship of breast density to breast cancer mortality risk**
A large study finding no difference in breast cancer mortality among women with dense breasts compared to others was discussed and referenced in the report.¹⁶
- 3. Discussion of the risk of overdiagnosis arising from supplemental screening**
Further discussion of the risk of overdiagnosis from supplemental screening was added. Our emphasis is on the need for longer term, comparative studies, which will help to quantify this risk.
- 4. Clarification to the effect of DBT on the recall rate**
The report was revised to clarify that existing evidence suggests that DBT reduces the recall rate.
- 5. Inclusion of additional studies of DBT**
All studies of DBT suggested by reviewers and in the public comment section were reviewed. None of these studies report stratified results for the study population of women with dense breasts, rather they report on outcomes of DBT in a general screening population of women. These studies were included in a separate brief report reviewing studies of the test performance characteristics of DBT in the general screening population.⁵³ Two studies mentioned in the public comments were added to this report.

6. Inclusion of data on the risk of radiation exposure from mammography for women with dense breasts

A separate analysis was conducted to estimate the risk of radiation-induced breast cancer arising from mammography.⁵⁴

USPSTF Involvement

This research was funded by the Agency for Healthcare Research and Quality (AHRQ) under a contract to support the work of the USPSTF. The authors worked with liaisons from the USPSTF at key points throughout the review process to develop and refine the scope, analytic framework, and key questions; to resolve issues around the review process; and to finalize the evidence synthesis. AHRQ had no role in study selection, quality assessment or synthesis. AHRQ staff provided project oversight, reviewed the draft evidence synthesis, and distributed the initial evidence report for external review of content by outside experts.

Chapter 3. Results

Literature Search

Our literature search yielded 2,067 unique citations. From these, we provisionally accepted 128 articles for review based on titles and abstracts (**Appendix A Figure 1**). After screening the full-text articles, 24 unique articles were determined to have met the inclusion criteria. The remaining 104 full-text articles were excluded (**Appendix C**).

We identified 5 studies (2 good-quality) that met inclusion criteria for KQ 1 (BI-RADS breast density determination reproducibility), 8 studies (5 good-quality) that met inclusion criteria for KQ 2 (test performance characteristics), 19 studies (5 good-quality) that met inclusion criteria for KQ 3 (proximate breast cancer outcomes), and 1 study (good-quality) that met inclusion criteria for KQ 4 (harms of density notification).

Key Question 1. What Is the Accuracy and Reproducibility of BI-RADS Determination of Breast Density?

Summary of Results

Since there is no recognized gold standard for breast density determination, we did not find studies that evaluated the accuracy of BI-RADS breast density determinations in screening mammography. We identified five studies that reported statistical measures for the reproducibility of categorical BI-RADS breast density classification among more than 440,000 women predominantly or exclusively receiving two sequential screening mammograms. We required that breast density categories in all included studies be based on the four BI-RADS categories, with or without percent density descriptions (**Table 1**), to reflect current U.S. practice. The majority of the evidence came from three studies set in the United States, with two based on data from the Breast Cancer Screening Consortium (BCSC)^{55, 56} and the other from a set of community radiologists conducting repeated readings of a large screening test set.⁵⁷ Two additional studies were based on much smaller samples or test sets from mammographic screening programs in Spain⁵⁸ or Italy.⁵⁹ All of the U.S.-based studies reflected community practice, by virtue of evaluating clinical readings from community screening programs or test set readings by practicing community radiologists without additional training.

Best estimates from these data suggest about one in five women would be categorized into a different BI-RADS density category (a, b, c, d) by the same radiologist after the second screening exam, while one in three would be categorized differently after a second exam read by a different radiologist. At a programmatic level, this translates into an estimated 12.6 to 18.7 percent of women being reclassified into a different overall combined breast density category (i.e. from “dense” [c or d] to “non-dense” [a or b] or vice-versa) after a second screening mammogram. These average estimates do not reflect some of the extremes seen when examining test-retest reproducibility among individual community-based radiologists. Breast density

categorizations based on sequential screening mammograms read by different radiologists showed lower overall agreement (kappas from 0.49 to 0.56, indicating moderate agreement) and a much greater range of inter-rater disagreements (kappas from 0.02 to 0.72) than seen with sequential readings by the same radiologist. Findings are most applicable to post-menopausal women or those aged 50 years and older, since these women comprised the bulk of those whose exams were studied.

Study Details

Two BCSC studies in the United States were in large population-based samples and represented community-assigned breast density readings between 1996 and 2009. The first good-quality study⁵⁵ evaluated breast density readings by a broad set of community radiologists of 871,502 digital or film-screen (FS) mammograms taken from 435,751 women aged 40 years and older, about a quarter of whom (28%) were 40 to 49 years of age. To be included, women had at least two screening mammograms conducted less than 36 months apart between January 1, 2000 and December 31, 2009.⁵⁵ The second fair-to-good-quality BCSC study evaluated the performance of 34 regional community radiologists who had interpreted at least 100 pairs of FS mammograms taken from 11,755 post-menopausal women with two or more screenings within a 3 to 24 month period between 1996 and 2006.⁵⁶ These two large community-based studies examining consistency of clinical readings were supplemented by a fair-quality study of 19 experienced community radiologists from a single U.S. integrated health care system who assigned BI-RADS breast density ratings to a test set of 341 FS mammograms on two occasions six-months apart in 2001 to 2002 in women of unspecified ages.⁵⁷ All three studies excluded women with history of breast cancer or augmentation, while the two BCSC studies also excluded women with recent hormonal medication use or other medical conditions that might introduce confounding due to physiologic changes in breast density between exams.^{55,56} The main risk of bias was due to unclear information about rater methods, including radiologist blinding to clinical information or other non-clinical cues.

Table 3 illustrates the relative consistency of breast density determinations for sequential examinations or readings at the population level. However, consistency in the distribution of BI-RADS density categorizations at the population level or measures of reproducibility (i.e., percent agreement, kappa) across all exams in a set can be misleading when considering individual patients. Even with the moderate or substantial agreement (kappas of 0.4 to 0.8) between readings, as seen across these large U.S. studies, 23 to 32 percent of women were assigned a different BI-RADS breast category in a sequential examination^{55,60} or when the community interpretation was compared to that of the majority of radiologists.⁵⁷

In the largest study,⁵⁵ the acquisition method (digital vs FS) overall, or in repeated exams (FS-FS, FS-digital, digital-FS, digital-digital), did not markedly affect the percentage of women for whom the radiologists assigned a different breast density from the first to second reading (i.e., increased or decreased at least one category). This allowed analyses to focus on the complete dataset, regardless of mammogram acquisition type, and reproducibility of categorical BI-RADS determinations by different radiologists between first and second examinations. Measures of agreement in BI-RADS categorization were high between readers of sequential mammograms (67% to 71%), and there was moderate overall interobserver agreement (kappa 0.49 to 0.56). The

lowest reproducibility of sequential BI-RADS breast density categorization occurred in the group initially assigned to the highest breast density category: BI-RADS d. In this relatively small group (7.5% of all women categorized as having “extremely dense” breasts on the first exam), just 57 percent were categorized the same after the second exam, with another 37 percent classified as “heterogeneously dense” (category c), 5 percent as “scattered fibroglandular densities” (category b), and under 1 percent as “almost entirely fatty” (category a) (data not shown). Although the majority of women (68.1%) were categorized in the same BI-RADS breast density category at both exams, almost one-third (31.9%) were categorized differently, with about one in five (18.7%) women reclassified into the opposite overall breast density category at their next screening exam (**Table 4**). Among the almost one-third of women categorized into a different BI-RADS density category on second exam (31.9%) (**Table 4**), breast density was reported in a lower category in 16.7 percent of women, and in a higher category in 14.7 percent of women (data not shown).

The probability of reclassification depended somewhat on the initial BI-RADS categorization. Of all those with a combined “dense breast” classification (c or d) on first exam, 21.7 percent were reclassified as “non-dense” (a or b) on second exam, while the percent reclassified from “non-dense” to “dense” was somewhat lower (16.3 percent). By far the largest proportion of reclassifications was between categories b and c. While these and other recategorizations primarily represented single category shifts, shifts between b and c are more critical since they represent a change in overall density categorization (i.e., between non-dense and dense). Overall, this study’s findings suggest that almost one in three women receiving breast density information would be told they were in different categories (i.e., a or b or c or d) after sequential screening exams less than three years apart. Considering the consequences of assigning women to different breast density categories according to legislative mandates, (i.e., from “dense” to “non-dense” or vice-versa), these data suggest that about 18.7 percent of women would receive discordant mandated notification suggesting different breast cancer risks and/or recommended actions within 36 months or less. Although changes in assigned breast density could reflect actual changes, these are less likely given the relatively short interval between exams (95% re-examined after fewer than 30 months) and control of medication use and other factors in the study population that would increase the likelihood of physiologic changes in breast density.

The second population-based BCSC study⁵⁶ focused on the reproducibility of individual community radiologists’ breast density readings among 11,755 post-menopausal women (mean age 66 years) receiving repeated FS mammographic screenings 3 to 24 months apart. Across all radiologists, percent agreement (77.2%, 95% CI 74.5 to 79.5) and intrarater reliability was moderate to substantial (kappa 0.58, 95% CI 0.55 to 0.61; weighted kappa 0.70, 95% CI 0.68 to 0.73). Reproducibility did not appear to vary by characteristics of the woman (i.e., age, change in BMI), by time between rescreening, or by changes in BI-RADS density definitions before and after the 4th edition (November 2003 to May 2004). However, individual radiologists demonstrated considerable variability in the reproducibility of their breast density readings in the same women, with 18 percent (6/34) having only slight or fair agreement, 59 percent (20/34) exhibiting moderate agreement, and 24 percent (8/34) with substantial agreement.

Although the distribution of women within breast density categories was quite similar for the first and second screening exams at the population level (**Table 3**), many individual women were

recategorized from first to second screening. The breast density of 22.8 percent of all women was categorized differently by the same radiologist at the second screening exam, with the highest proportion of women receiving a different density rating among those categorized initially in the lowest and highest breast density categories (i.e., 45% and 50% respectively). Considering potential actions based on the first exam, 19.1 percent of women categorized as having dense breasts initially (c or d) would not be notified of dense breasts with their second exam results (a or b). Similarly, 10 percent not classified as having dense breasts on first exam (a or b) would have been notified that they had dense breasts (c or d) after the second exam. Given the distribution of results (i.e. most women were assigned breast density b or c initially), the largest number of reclassifications were between categories b and c, which are single category changes but represent an important change in the overall density categorization. This study's results suggest that, overall, about one in five women would be categorized into a different BI-RADS breast density category in a sequential screening mammogram by the same radiologist; 12.6 percent of all women screened would potentially be advised differently according to legislative mandates (**Table 4**).

A fair-quality U.S. study examined both inter-rater and intra-rater reliability among 19 experienced radiologists from a single integrated U.S. health care system. The study compared repeated readings of a test set of FS mammograms six months apart (2001 to 2002); films represented screening exams in 341 women from the same institution.⁵⁷ BI-RADS density determinations for each woman's films were compared between pairs of radiologists (and between repeated readings for each radiologist). Inter-rater agreement was fair (mean kappa 0.46; 95% CI, 0.36 to 0.55) and varied widely between pairs of radiologists, from slight to substantial; one radiologist displayed only slight agreement with any other. Agreement between radiologists did not vary by radiologist experience, time spent reading mammograms/breast imaging, or findings (cancer versus not). Intra-rater agreement was higher, with substantial mean agreement (mean kappa 0.69, 95% CI 0.63 to 0.73) and no less than moderate agreement for repeated breast density readings by any individual radiologist; radiologists with more than 10 years of experience interpreting mammograms had higher reproducibility of their own breast density readings (increased mean kappa 0.10; 95% CI, 0.01 to 0.24).

When the original clinical breast density readings were compared to the majority opinion of all readers, 29 percent of women were categorized into a different BI-RADS density category (a, b, c, d) (**Table 4**). As with other studies, the majority of exams reclassified from "dense" to "non-dense" represented changes in categorization from b to c (or vice versa), but those initially designated to the highest and lowest categories a or d by the original clinical readers were the least stable. Although many (55%) of the BI-RADS category a classifications assigned during the original clinical readings were re-rated category b by the majority of radiologists, none were assigned categories c or d. Similarly, many (40%) of the category d original clinical readings were reassigned category c by the majority of radiologists, but none moved to a "non-dense" (a or b) group. In contrast, 25.9 percent of films originally categorized as b on the clinical reading were re-categorized as c by the majority of readers, with a smaller percentage (13.4%) being re-categorized from c to b by the majority. Based on these re-categorizations, an estimated 17 percent of screened women would have been reassigned to the opposite breast density category ("dense" vs "non-dense" by the majority of radiologists).

This study illustrates the high potential for misclassification when breast density is assigned by single readers and there are outlier performers, as occurred among these 19 radiologists. Among 145 examinations classified as “non-dense” (a or b) by the majority of readers, 83 percent were interpreted as “dense” (c or d) by at least one radiologist. Similarly, at least one radiologist interpreted almost half (47%) of the 187 examinations interpreted as “dense” by the majority as “non-dense.” Almost all (93%, n=316) exams were interpreted as “dense” (c or d) by at least one radiologist. These data illustrate the clear potential for misclassification of breast density after a single reading by a community radiologist.

Two studies outside the United States examined the reproducibility of BI-RADS breast density classifications among experienced radiologists reading test sets of 100 digital or FS mammograms from screening programs in Spain and Italy (data not shown in tables). Both programs provided initial training to calibrate performance among the 21 experienced radiologists⁵⁸ or to orient six otherwise experienced radiologists not currently using the BI-RADS system for breast density.⁵⁹ After calibration, a good-quality study of 21 radiologists found moderate to substantial inter-rater agreement across all four breast density categories (kappa 0.44; weighted kappa 0.73); intra-rater agreement was higher (kappa 0.64; weighted kappa 0.82). Nonetheless, among pairs of radiologists, 15.1 percent of mammograms (3,180/20,694) classified as “non-dense” by one reader were classified as “dense” by another. Within individual radiologists, repeat classification of the same image between non-dense and dense categorization was less often divergent (i.e., 10.4%). Among six Italian radiologists newly oriented to using the BI-RADS four categories for breast density classification, agreement between pairs of radiologists across all four categories was substantial or better (weighted kappas ranging from 0.61 to 0.87); agreement was slightly improved across collapsed categories (“non-dense” vs “dense”) with weighted kappas ranging from 0.64 to 0.94. Nonetheless, depending on radiologist pairs, 6 to 15 percent of women were allocated differently to breast density categories after different single-readers.

Key Question 2. What Are the Test Performance Characteristics of Newer Technologies for Breast Cancer Screening When Used as Supplemental Tests After a Negative Screening Mammography Exam in Women Found to Have Dense Breasts? How Do These Performance Characteristics Differ by Age and Risk Factors?

Summary of Results

We identified and reviewed studies of diagnostic test characteristics (with defined reference standards) for hand held breast ultrasound (HHUS), automated whole breast ultrasound (ABUS), and breast magnetic resonance imaging (MRI) among women with dense breasts and negative screening mammography. We did not identify any studies of the diagnostic test characteristics of digital breast tomosynthesis (DBT) for women with dense breasts, which is a relatively new technology with limited research. We excluded studies that did not followup negative test results

for a minimum of one year as part of the reference standard. Most included studies had cohort designs, assessing the accuracy of supplemental test results within a group of women undergoing supplemental screening in addition to screening mammography. When possible, we assessed the performance of supplemental tests among women with negative screening mammography. In these studies, the reference standard was either tissue sampling for women recommended for biopsy or a minimum of one year clinical followup for women with either negative supplemental screening results or who were not recommended for biopsy. Incremental sensitivity was calculated based on cancers detected by the supplemental screening test and interval cancers. Cancers detected by mammography were not included in the calculation of sensitivity, as the focus was on supplemental testing in women found to have dense breasts after a negative mammogram.

Two good-quality studies (one from the United States⁶¹ and one from Italy⁶²) and three fair-quality studies (one each from the United States,⁶³ Singapore⁶⁴ and South Korea⁶⁵) reported on HHUS (**Tables 5 and 9**); one fair-quality study from the United States reported on ABUS (**Tables 6 and 10**);^{66, 67} three good-quality studies (one each from the United States,⁶¹ Germany,²³ and the Netherlands⁶⁸) reported on MRI (**Tables 7 and 11**). Some studies focused exclusively on women with BI-RADS density c to d, while others reported on mixed populations including women with elevated risk due to BRCA 1/2 mutations, a personal history of breast cancer, or family history of breast cancer. Of these, we included only studies in which a majority of women had dense breasts, and in some cases we were able to isolate a subgroup with dense breasts without a personal history of breast cancer or BRCA 1/2 mutation carriage. In two good-quality studies,^{23, 61} the authors provided unpublished data for the subgroup of women with dense breasts, and removed the women with BRCA 1/2 mutations, histories of chest irradiation, and/or personal histories of breast cancer from the subgroup. We found no studies reporting how the performance of these modalities varies by patient age and other risk factors for breast cancer among women with dense breasts.

In two good-quality studies, the sensitivity of HHUS for detecting any breast cancer (including invasive breast cancer and DCIS) among women with dense breasts after recent negative screening mammography ranged from 80.0 percent (95% CI, 64.5 to 91.0)⁶² to 83.3 percent (95% CI, 58.6 to 96.4)⁶¹ and for invasive breast cancer, the sensitivity ranged from 77.8 percent (95% CI, 60.9 to 89.9)⁶² to 82.4 percent⁶¹ (95% CI, 56.6 to 96.2) (**Table 13; Figures 2-5**). Specificity ranged from 86.4 percent (95% CI, 85.2 to 87.5 percent)⁶¹ to 94.5 percent (95% CI, 94.0 to 95 percent)⁶² for all breast cancer and from 86.4 percent (95% CI, 85.2 to 87.5)⁶¹ to 94.5 percent (95% CI, 94.0 to 95.0)⁶² for invasive breast cancer. Positive predictive value (PPV) ranged from 3.0 percent⁶¹ to 6.6 percent⁶² for invasive cancer. Negative predictive value (NPV) was uniformly high, ranging from 99.8 percent⁶¹ to 99.9 percent.⁶²

One fair-quality study from the United States reported on screening women (68% had dense breasts) with ABUS.⁶⁶ For the entire group, the sensitivity of ABUS after negative mammography was 67.6 percent (95% CI, 49.5 to 82.6%), specificity was 91.6 percent (95% CI, 91.0 to 92.3%), PPV was 4.1 percent and NPV was 99.8 percent. For invasive cancer only, sensitivity was 66.7 percent (95% CI, 48.2 to 82.0%), specificity was 91.6 percent (95% CI, 90.9 to 92.3%), PPV was 3.9 percent, and NPV was 99.8 percent (**Table 14; Figures 2-5**). The study's authors had a major financial interest in the ABUS system being evaluated.

Three good-quality studies evaluated supplemental MRI screening among women with dense breasts and a negative screening mammogram. The sensitivity of supplemental MRI screening for all breast cancers (including invasive breast cancer and DCIS) was 100 percent in two good-quality studies^{23, 61} and 75 percent (95% CI, 34.9 to 96.8%) in a third good-quality study⁶⁸ (**Table 15; Figures 2-5**). In two studies,^{23, 61} the authors provided us with unpublished data for the subgroup of women with dense breasts, excluding women with BRCA 1/2 mutations, histories of chest irradiation, or personal histories of breast cancer. In both of these studies, women had also had recent negative screening with HHUS. Sensitivity of supplemental MRI screening was high in all three studies, but the numbers of incident breast cancers in the analyzed subgroups were small (2 to 8), so sensitivity estimates have substantial uncertainty. In all three studies, most of the cancers detected by MRI were invasive rather than DCIS. Specificity ranged from 78.1 percent (95% CI, 73.2 to 82.5)⁶¹ to 93.2 percent²³ (95% CI, 86.5 to 97.2), while the PPV of an abnormal MRI result ranged from 3.0 percent⁶⁸ to 22.2 percent.²³ Negative predictive values were very high in all three studies (99.9-100%).^{23, 61, 68}

In summary, most studies of diagnostic test performance of supplemental screening tests for women with dense breasts included women selected for elevated breast cancer risk, though women with BRCA 1/2 mutations, chest wall irradiation, and a personal history of breast cancer were excluded for the subgroups we analyzed. In these study subgroups, the sensitivity of supplemental MRI screening after negative screening mammography is likely to be higher than HHUS screening, but few studies to date have compared sensitivity of these screening modalities among women with dense breasts. Specificity of these modalities is similar, and PPV is low. No summary is possible for ABUS and DBT, with only a single study of ABUS and no studies of DBT identified for review. We found no studies of any supplemental test that examined the effect of age and other breast cancer risk factors on diagnostic test performance characteristics in women with dense breasts.

Study Details

HHUS

One good-quality prospective diagnostic accuracy study from the United States reported on one to three rounds of HHUS screening at 12 month intervals following negative mammography in women with at least one breast quadrant identified as dense and at least one other risk factor for breast cancer.⁶¹ In this study, 7,473 exams were performed on 2,662 women (mean age 55.2 years). The reference standard was defined as the most severe biopsy results or clinical followup for 1 year. Loss to followup was less than 3 percent. Because 53 percent of participants had a personal history of breast cancer, the study authors provided us with supplemental unpublished data on the subgroup of 1,216 women undergoing 3,414 exams after excluding BRCA 1/2 mutation carriers, those with a history of breast cancer, or prior chest, mediastinal or axillary irradiation. In this subgroup, three interval cancers were identified on clinical followup. The incremental sensitivity of HHUS following negative mammography for all breast cancer was 83.3 percent (95% CI, 58.6 to 96.4), while specificity was 86.4 percent (95% CI, 85.2 to 87.5), PPV was 3.0 percent, and NPV was 99.0 percent. For invasive cancer, sensitivity was 82.4 percent (95% CI, 56.6 to 96.2), specificity was 86.4% (95% CI, 85.2 to 87.5), PPV was 3.0

percent, and NPV was 99.9 percent. Of note, in this subgroup, 334 women underwent MRI screening after 3 negative mammograms and 3 negative HHUS exams over a 24-month period, and an additional 7 cancers (6 invasive) were identified (see MRI section for additional details).

A good-quality retrospective study from an Italian charity breast screening clinic included 3,556 women with BI-RADS c to d density undergoing 7,224 HHUS screening exams following a negative screening mammogram.⁶² Followup for interval cancers included searching clinic archives and data linkage with hospital discharge records; the authors estimated the probability that interval cancers were not identified was less than 5 percent. Eight interval cancers were identified. The incremental sensitivity of HHUS following negative mammography was 80 percent (95% CI, 64.5 to 91.0), specificity was 94.5 percent (95% CI, 94.0 to 95.0), PPV was 7.5 percent and NPV was 99 percent. For invasive cancer, sensitivity was 77.8 percent (95% CI, 60.9 to 89.9%), specificity was 94.5 percent (95% CI, 94.0 to 95.0), PPV was 6.6% and NPV was 99.9%.

A smaller fair-quality diagnostic accuracy study from Connecticut reported on 935 women with heterogeneous or extremely dense breasts undergoing a single round of HHUS, of whom 648 had a recent negative screening mammogram.⁶³ HHUS was performed at a university radiology practice by mammography technologists and read by one of eight breast radiologists who were aware of the prior mammography results. Followup was based on biopsy results or mammography and/or HHUS performed at least 15 months after the initial exam. Nineteen percent were lost to followup. No interval cancers were detected. Among the 648 women with negative mammograms, for all breast cancer, sensitivity was 100 percent (95% CI, 29.2 to 100), specificity was 76.7 percent (95% CI, 73.3 to 79.8), PPV was 2.0 percent and NPV was 100 percent. For invasive breast cancer, sensitivity was 100 percent (95% CI, 15.8 to 100), specificity was 76.6 percent (95% CI, 73.2 to 79.8), PPV was 1.3 percent, and NPV was 100 percent.

Women of Asian ethnicity have been noted to have a higher proportion of BI-RADS c or d breast density in the United States.^{19, 20} Two fair-quality studies from Asia^{64, 65} addressed diagnostic test characteristics for women with dense breasts. In a study of 1,046 women with dense breasts undergoing 1,507 HHUS exams at an academic medical center in South Korea, 61.8 percent of participants had a personal history of breast cancer; however outcomes were reported separately for women without a history of breast cancer (446 screening exams).⁶⁵ Clinical followup extended over 2 years; no interval cancers were identified. In the subgroup of women without a personal history of breast cancer, the incremental sensitivity of screening HHUS was 100 percent (95% CI, 71.5 to 100), specificity was 71.7 percent (95% CI, 67.2 to 75.9), PPV was 8.2 percent and NPV was 100 percent. Breast cancers were not stratified as invasive versus DCIS.

A small study from a general hospital radiology practice in Singapore reported screening HHUS on 141 women with negative mammograms; all had heterogeneous or extremely dense breasts and 5 percent had a personal history of breast cancer.⁶⁴ Followup consisted of a retrospective review of all breast imaging at two years; 35 women (25%) were lost to followup, leaving 106 women for assessment of diagnostic test characteristics. No interval cancers were identified. For all breast cancer, sensitivity of HHUS after negative mammography was 100 percent (95% CI, 15.8 to 100%), specificity was 78.9 percent (95% CI, 69.7 to 86.2), PPV was 8.3 percent and NPV was 100 percent. For invasive breast cancer, sensitivity was 100 percent (95% CI, 2.5 to

100), specificity was 80 percent (95% CI, 71.4 to 86.5), PPV was 4.5 percent, and NPV was 100 percent.

ABUS

One fair-quality diagnostic test performance study from the United States reported screening women with ABUS.⁶⁶ An important limitation of the study was that the lead author on the study had a majority financial interest in the ABUS technology being studied, and the co-author was also a shareholder. This study reported on a group of 4,419 women undergoing 6,425 ABUS and mammography exams, of which 4,347 exams (67.7%) were performed on women with BI-RADS c to d breast density. Ten percent of participants had a personal history of breast cancer and 30 percent had a family history of breast cancer; 0.1 percent were known BRCA 1/2 mutation carriers. Over one year, 20 percent of the group was lost to followup. For the entire group, ABUS sensitivity for all breast cancer in women with negative mammography was 67.6 percent (95% CI, 49.5 to 82.6), specificity was 91.6 percent (95% CI, 91.0 to 92.3), PPV was 4.1 percent and NPV was 98.8 percent. For invasive breast cancer, sensitivity was 66.7 percent (95% CI, 48.2 to 82.0), specificity was 91.6 percent (95% CI, 90.9 to 92.3), PPV was 3.9 percent and NPV was 99.8 percent.

MRI

The largest sample of women with dense breasts receiving supplemental MRI screening derived from the good-quality, multi-center Dutch MRISC study.⁶⁸ Conducted from 1999 to 2003, MRISC recruited asymptomatic women aged 25 to 70 years who had a lifetime breast cancer risk of ≥ 15 percent to undergo annual MRI screening in addition to annual mammography screening and screening clinical breast examination every 6 months (mean 2.3 screening rounds). All of the women had a family history of breast cancer. For this review, we abstracted data for a subgroup of women with BI-RADS c or d breast density without BRCA 1/2 gene mutations who underwent 1,723 exams. Among these women, breast cancer was diagnosed in 8 women, and 6 of these cancers were detected by MRI (sensitivity 75% [95% CI, 34.9 to 96.8]). Nearly all women (1,522 of 1,524) with negative MRI screening results remained breast cancer free during the following year. The specificity of screening MRI was 88.8 percent (95% CI, 87.2 to 90.2), PPV was 3.0 percent and NPV was 99.9 percent. In the subgroup of women with dense breasts, sensitivity of supplemental MRI for invasive cancer vs. DCIS was not reported.

Authors of a good-quality, multi-center U.S. study provided us with unpublished subgroup data for 334 asymptomatic women with elevated breast cancer risk and heterogeneously or extremely dense breasts (without BRCA 1/2 mutation, personal history of breast cancer, or history of chest wall irradiation) who underwent a single round of MRI screening following three annual rounds of mammographic and breast ultrasound screening.⁶¹ For these women, the most recent mammogram and breast ultrasound were interpreted as normal. Of the 334 women, 7 were diagnosed with breast cancer by MRI screening, and there were no interval cancers in one year of followup, resulting in a sensitivity of 100 percent (95% CI, 59.0 to 100). The screening MRI was read as abnormal in 78 of 334 women, yielding a specificity of 78.1 percent (95% CI, 73.2 to 82.5) and a PPV of 9.0 percent; NPV was 100 percent. Six of the seven cancers were invasive; sensitivity for invasive cancer was 100 percent (95% CI, 54.1 to 100), specificity was 78.8%

(95% CI, 72.9 to 82.2), PPV was 7.7 percent and NPV was 100 percent.

In a good-quality single-center German study, a subgroup of 105 asymptomatic women with dense breasts underwent one or two rounds of annual MRI screening following negative screening mammography and screening ultrasound.²³ An aim of the study was to demonstrate comparable interpretive results with abbreviated MRI protocol versus a full diagnostic protocol. We report outcomes using the full diagnostic protocol. In the 105 women, 3 breast cancers were diagnosed within two years of screening (2 invasive cancers and 1 DCIS), all of which were MRI detected, yielding a sensitivity of 100 percent (95% CI, 29.2 to 100). There were no interval breast cancers during two years of followup, so the NPV was 100 percent. MRI was read as abnormal in 9 women, yielding a specificity of 94.1 percent (95% CI, 87.6 to 97.8) and a PPV of 33.3 percent. For invasive breast cancer, sensitivity was 100 percent (95% CI, 15.8 to 100%), specificity was 93.2 percent (95% CI, 86.5 to 97.2), PPV was 22.2 percent and NPV was 99.9 percent. Performance was similar with the abbreviated and full diagnostic protocols.

It is notable that the two of three studies on supplemental MRI screening in women with dense breasts targeted women with elevated lifetime breast cancer risk.^{61, 68} Thus, the subgroup of women with dense breasts in these studies likely had elevated breast cancer risk based on risk factors in addition to breast density (e.g., family history of breast cancer). Only one study included a subgroup that received supplemental MRI screening solely on account of breast density.²³

DBT

We found no studies of the test performance characteristics of DBT among women with dense breasts.

Key Question 3. When Performed After a Negative Screening Mammogram in Women Found to Have Dense Breasts, What Is the Effectiveness of Supplemental Screening With Breast Ultrasound, MRI, or DBT on Proximate Clinical Outcomes, Including Cancer Detection Rates, DCIS Detection Rates, Stage at Diagnosis, Recall Rates, Biopsy Rates, and Interval Cancer Rates?

Summary of Results

We found no RCTs comparing any clinical outcomes with and without supplemental screening. We tended toward broad inclusion of observational studies reporting on clinical outcomes of supplemental screening women with dense breasts to identify as much information as possible, and considered studies reporting breast density based on the BI-RADS density system, Wolf classification (>50% density) or selecting women described as having heterogeneously or extremely dense breasts (**Tables 5-8**). We identified ten studies of HHUS^{61-65, 69-73} (two good-

quality^{61, 62}), three fair-quality studies of ABUS,^{57, 66, 67, 74} three good-quality studies of MRI,^{23, 61, 68} and four fair-quality studies of DBT.⁷⁵⁻⁷⁸ Many studies selected women at higher risk for breast cancer due to risk factors other than increased breast density, such as BRCA 1/2 mutation carriage or personal history of breast cancer. We included studies if we were able to abstract a subgroup with dense breasts on mammography or if factors conferring very high risk for breast cancer were not highly prevalent. In the case of two good-quality studies,^{23, 61} the authors provided us unpublished data on the study subgroup with dense breasts and excluded women at high risk due to BRCA 1/2 mutations, history of chest wall irradiation, or a personal history of breast cancer. Most studies reported descriptively on incremental impacts on outcomes after supplemental testing among a cohort of women with dense breasts who had recent negative mammography. Two studies of HHUS,^{61, 62} two studies of ABUS,^{66, 74} three studies of MRI^{23, 61, 68} and one of DBT⁷⁸ compared proximate outcomes from mammography to outcomes in the same group with supplemental testing. One fair-quality study of ABUS⁶⁷ and three fair-quality studies of DBT compared proximate outcomes among two groups of women, one undergoing mammography and another undergoing mammography as well as supplemental testing. Only one of these studies adjusted for differences between the groups.⁷⁵

In general, supplemental testing consistently detected additional breast cancers not identified by mammography (**Tables 16-20; Figure 6**). Findings from good-quality studies for supplemental testing modalities are found in **Table 20**. The two good-quality studies for HHUS (one each from the U.S. and Italy) were consistent in their estimates of the incremental cancer detection rate of 4.4 (95% CI, 2.5 to 7.2) per 1,000 exams,^{61, 62} In these two studies, the majority of cancers were invasive: 14/15 (93%)⁶¹ and 28/32 (88%).⁶² In the same groups of women, mammography cancer detection rates were 4.7⁶¹ in the U.S. study and 2.8⁶² per 1,000 exams in the Italian study. In three good-quality studies of women with dense breasts undergoing supplemental MRI screening, incremental breast cancer detection rates ranged from 3.5 per 1,000 exams (95% CI, 1.3 to 7.6)⁶⁸ to 21.0 per 1,000 exams (95% CI, 8.5 to 42.7).⁶¹ Mammography cancer detection rates in two of these studies for women with dense breasts were 4.1⁶⁸ and 7.0⁶¹ per 1,000 exams. In studies of supplemental MRI screening, MRI detected predominantly small, early-stage invasive breast cancers. However, the overall numbers of detected cancers were small (2 to 7), and women in these studies had higher lifetime breast cancer risk than the general population of women with dense breasts.

Recall rates for supplemental HHUS (after negative mammography) varied in these good quality studies and were reported only in a single U.S. study: 13.9 percent.⁶¹ Biopsy rates were 6.9 percent in the in the U.S. study⁶¹ compared to 5.9 percent in the Italian study.⁶² In the good-quality studies of supplemental MRI screening in women with dense breasts, results were interpreted as abnormal at rates ranging from 8.6 percent⁶⁸ to 23.4 percent.⁶¹ For two studies, data on biopsy rates in subgroups of women with dense breasts in the MRI studies were either not reported⁶⁸ or were not available.^{23, 61} However, in a good-quality study in the United States, 7.0 percent of entire population of women undergoing supplemental MRI screening underwent biopsy on account of abnormal MRI results (the incremental biopsy rate) following negative screening mammography and ultrasound.⁶¹ Because most studies reported on only 1 (at most 3) rounds of screening, the cumulative effect of recall for additional imaging and biopsy would be greater over time, with additional rounds of screening.

Only fair-quality studies were identified for ABUS and DBT. For ABUS the incremental cancer detection rate ranged from 1.8 per 1,000 exams (95% CI 1.2 to 2.6)⁷⁴ to 15.1 per 1,000 exams (95% CI, 11.4 to 19.9).⁶⁷ In one of these studies, the cancer detection rate from mammography was 4.7 per 1000 exams (95% CI, 2.8 to 7.3).⁶⁶

Four fair-quality studies of DBT reported separately on screening populations of women with dense breasts.⁷⁵⁻⁷⁸ Three of these studies reported on cancer detection rates: detection rates with digital mammography in women with dense breasts were similar (4.0 to 5.2 per 1,000 exams) and were higher with the addition of DBT (5.4 per 1,000 exams [95% CI, 3.5 to 7.9]⁷⁷ to 6.9 per 1,000 exams [95% CI, 4.8 to 9.6]⁷⁵) (**Table 19**). Invasive cancer detection rates for women with dense breasts were reported only for one study: 3.4 per 1,000 exams with digital mammography and 4.7 per 1,000 exams with the addition of DBT.⁷⁵ Recall rates were reported by all four studies, though recalls for additional imaging were not distinguished from recalls for biopsy. Recall rates declined with the addition of DBT. Recall rates were higher in the U.S. studies, ranging from 9.1 percent⁷⁷ to 16.6 percent⁷⁶ for digital mammography compared to 6.9 percent⁷⁷ to 10.8 percent⁷⁵ with the addition of DBT. In the Italian study, in which all digital mammograms and DBT images underwent double-reading (i.e., independent interpretation by two radiologists), recall rates declined from 7.2 percent to 6.6 percent with the addition of DBT.⁷⁰ Biopsy rates were not reported separately for women with dense breasts in any of these studies.

Overall, the evidence indicates that supplemental screening among women with dense breasts and recent negative mammography can consistently identify additional cancers, most of which are invasive, but also leads to additional recalls and biopsies, with the possible exception of DBT, though evidence is quite limited. Studies of proximate screening outcomes among women with dense breasts undergoing supplemental screening tests were predominately designed to estimate the incremental impact of supplemental testing on cancer detection rates and diagnostic testing. Many lacked sufficient followup to identify false negatives, and none compared interval breast cancer rates or potential surrogates for breast cancer mortality among two groups of women with dense breasts undergoing screening mammography with vs. without supplemental testing. Such studies would potentially provide greater insight regarding long-term benefits and harms of supplemental screening. In addition, many study populations were at higher breast cancer risk than would be conferred by increased breast density alone, raising questions about the generalizability of findings to the broad population of women with dense breasts. Variability in breast cancer risk may explain some observed differences in rates of breast cancer detection by mammography and supplemental testing. Other sources of variation may relate to variability in skill and experience among interpreting radiologists, and variation in technology used for supplemental testing. Long terms RCTs or well-designed comparative observational studies could provide much stronger evidence about meaningful outcomes of supplemental screening for women with dense breasts.

Study Details

HHUS

One good-quality prospective diagnostic accuracy study from the United States reported on one

to three rounds of HHUS screening at 12 month intervals in women with dense breasts and at least one other risk factor for breast cancer.⁶¹ In this study, 7,473 exams were performed on 2,662 women (mean age, 55.2 years). Because 53 percent of participants had a personal history of breast cancer, the authors provided unpublished data on the subgroup of 1,216 women undergoing 3,414 exams, excluding BRCA 1/2 mutation carriers and those with a history of breast cancer chest irradiation. Loss to followup was less than 3 percent. In this subgroup, screening mammography detected 24 cancers (cancer detection rate, 19.7 per 1,000 exams), 15 of which were invasive. Among women with negative mammography, HHUS after negative mammography detected an additional 15 breast cancers (14 invasive, 1 DCIS), for a cancer detection rate of 4.4 per 1,000 exams (95% CI, 2.5 to 7.2). Three interval cancers were identified on clinical followup. The stage distribution for the subgroup was not available. In the overall group for those cancers where nodal staging was available, 1/27 cancers detected by HHUS had positive axillary nodes. The subgroup recall rate was 13.9 percent (95% CI, 12.9 to 14.0) for HHUS. Biopsy rates were not available for the subgroup. In the overall group, 8.8 percent of women underwent biopsy on account of HHUS results during the initial screen, and during subsequent screening rounds, the biopsy rate due to HHUS was 5.5 percent.

A good-quality diagnostic accuracy study from an Italian breast cancer screening program reported on 3,356 asymptomatic women with BI-RADS c to d density undergoing 7,224 HHUS screening exams following a negative screening mammogram.⁶² Fifty five percent were less than 50 years of age. Based on searches of clinic archives and data linkage with hospital discharge records, the authors estimated the probability that interval cancers were not identified was less than 5 percent. Screening HHUS detected 32 cancers, for an incremental cancer detection rate of 4.4 per 1,000 exams (95% CI, 3.0 to 6.2). Twenty eight of these were invasive (cancer detection rate, 3.9 per 1,000 exams [95% CI, 2.6 to 5.6]). Of the 32 cancers, 4 were in-situ, 24 were Stage I, 3 were Stage II, and one was unknown stage. Mammography in the same group detected 20 breast cancers (cancer detection rate, 2.8 per 1,000 exams); 4 in-situ, 11 Stage I, 1 Stage II, and 4 were unknown stage. There were 8 interval cancers identified within 1 year of screening, 2 were in-situ and 6 were Stage I. The total recall rate was not reported, but the biopsy rate was 5.9 percent (95% CI, 5.4 to 6.5).⁶²

Three fair-quality studies reported on experience with HHUS in Connecticut, where insurance coverage of ultrasound screening for women with dense breasts was mandated in 2005, and breast density reporting to women as part of mammography results was mandated as of October 2009.^{63, 71, 73} A smaller fair-quality diagnostic accuracy study from Connecticut reported on 935 women (mean age, 52 years) with heterogeneous or extremely dense breasts undergoing a single round of HHUS, of whom 648 had a recent negative screening mammogram.⁶³ The incremental cancer detection rate of HHUS was 3/638 or 4.6 per 1,000 exams (95% CI, 1.0 to 13.5). Two of the cancers were invasive and no interval cancers were detected. The overall recall rate was 23.6 percent and the biopsy rate was 7.0 percent. One cohort study reported screening results for a single screening round of 8,647 women undergoing HHUS after a negative screening mammogram at one of 6 radiology practices (mean age, 54.4 years).⁷³ HHUS detected 25 breast cancers (21 invasive, 4 DCIS) among the 8,647 exams with negative mammograms for a cancer detection rate of 2.9 per 1,000 exams (95% CI, 1.9 to 4.3). The overall recall rate was 13.8 percent, with a biopsy rate of 4.8 percent. No comprehensive followup was reported. A single center study reported on findings from 5,519 women (mean age, 53.6 years) undergoing a single

screening HHUS following mammography after the mandated reporting law went into effect.⁷¹ The study included HHUS findings from some women who had abnormal mammograms but reported only HHUS results from the breast quadrants noted as normal on mammography (the number of these exams included was not recorded). Eighty nine percent had heterogeneous or very dense breasts, and 6 percent had a personal history of breast cancer. HHUS detected 10 breast cancers (9 invasive, 1 DCIS) for a cancer detection rate of 1.8 per 1,000 exams (95% CI, 0.9 to 3.3). The overall biopsy rate for HHUS (biopsy was recommended for suspicious or highly suspicious findings – BI-RADS 4-5) was 3.2 percent (95% CI, 2.8 to 3.8). The number of recommendations for short interval followup was not reported.

Three fair-quality studies from Italy reported on clinical outcomes of HHUS screening in women with dense breasts. In a smaller cohort study of fair-quality, 1,666 women aged 40 to 49 years with an average lifetime risk of 11.6 percent were recruited, of whom 800 underwent one round of screening breast HHUS for BI-RADS c/d breast density, and 26 for breast implants.⁷² This combined group of 826 had an incremental cancer detection rate of 2.4 per 1,000 exams (95% CI, 0.3 to 8.7) and a recall rate of 9.5 percent (95% CI, 7.5 to 11.6). Ten cancers were detected among the 826 women on mammography; two additional cancers were diagnosed on HHUS, one Stage 1 and one Stage 2. No followup of negative HHUS screens was reported. A study of 5,227 women (68% were aged 40 to 49 years) with BI-RADS density c to d and negative mammography were screened with HHUS within 1 month of the mammogram.⁶⁹ Two cancers (both invasive) were identified by HHUS, for a cancer detection rate of 0.4 per 1,000 exams (95% CI, 0.0 to 1.4). One cancer was Stage 1, and one was Stage 2. The recall rate was 2.1 percent (95% CI, 1.7 to 2.5) and the biopsy rate was 1.2 percent (95% CI, 1.0 to 1.6). A cohort study from a breast clinic reported on findings of 22,131 women undergoing breast HHUS after negative mammography, of whom 9,960 had BI-RADS c to d dense breasts.⁷⁰ The cancer detection rate in this group with dense breasts was 2.2 per 1,000 exams (95% CI, 1.4 to 3.3), somewhat higher than the rate of 1.56 per 1,000 reported for women with BI-RADS a to b breast density. No other findings were reported separately for the subgroup with dense breasts.

Two fair-quality retrospective diagnostic accuracy studies from Asia reported on proximate clinical outcomes from HHUS screening of women with dense breasts who had negative mammography.^{64, 65} These studies may be potentially relevant to women of Asian descent residing in the U.S., who have greater risk of increased breast density.¹⁹⁻²¹ An academic medical center in South Korea reported outcomes for 1,507 exams on 1,046 women (mean age 47.5 years) with two years followup but 61.8 percent had a personal history of breast cancer and others were undergoing a diagnostic exam; 446 HHUS exams were for screening of women with heterogeneous or extremely dense breasts without a personal history of breast cancer.⁶⁵ In this subgroup, 11 breast cancers were detected, for a very high cancer detection rate of 24.7 per 1,000 exams (95% CI, 12.4 to 43.7), but study estimates may be biased by loss to followup, since the exams were selected based on availability of two year followup and 60.5 percent (2,313/3,820 exams) originally identified were initially excluded on this basis. The stage distribution was not clearly reported for the subgroup. The recall rate was 13.7 percent (95% CI, 10.6 to 17.2) with a biopsy rate of 10.9 percent (95% CI, 8.2 to 14.3). A Singapore study of 141 women (mean age 45.1 years) with heterogeneously or extremely dense breasts and negative mammograms reported on a single round of HHUS screening at a general hospital radiology practice.⁶⁴ Followup was by retrospective review of breast imaging records over 24 months following

screening: 35 women (25%) were lost to followup, leaving 106 women for analysis. Two breast cancers (one invasive- Stage 1) were diagnosed for an incremental cancer detection rate of 18.9 per 1,000 women (95% CI, 1.7 to 50.3). No interval cancers were reported. The recall rate was 17 percent, and the biopsy rate was 13.2 percent (95% CI, 7.4 to 21.2).

ABUS

Three fair-quality cohort studies from the United States reported proximate outcomes of screening women with ABUS. One study from an academic medical center compared proximate outcomes among women with >50 percent density (Wolf classification) on recent mammography: 4,076 were screened with mammography alone and 3,418 were women screened during a later period with both mammography and ABUS.⁶⁷ All imaging was double-read by two radiologists who interpreted all mammograms and ABUS exams with consensus over disagreements. One round of screening results was reported, but no description of the followup protocol, loss to followup, or interval cancers was provided. ABUS with mammography identified 12.3 invasive breast cancers per 1,000 exams (95% CI, 8.9 to 16.6) relative to 4.7 invasive breast cancers per 1,000 exams identified in the mammography only group. The recall rate in the ABUS with mammography group was 1.5 percent (95% CI, 1.1 to 2.0); the biopsy rate for additional imaging was not reported. No data were provided on the number of short interval followup exams recommended. Breast cancer stage was reported only for the group undergoing both ABUS and mammography. Forty two breast cancers were diagnosed by ABUS, and all were invasive. Thirty five (83%) were Stage I, five (12.5%) were Stage II and two (5%) were Stage III. The interval cancer rate was not reported, though the authors state that patients were followed prospectively for 1 year. Loss to followup was not reported.

A fair-quality study of test performance characteristics from eight radiology breast screening facilities across the United States reported on a group of 4,419 women (mean age, 53 years) undergoing 6,425 ABUS and mammography exams, of which 4,347 exams (67.7%) were performed on women with BI-RADS c to d breast density.⁶⁶ Most women underwent concurrent mammography and ABUS, but 468 women alternated mammography with ABUS at six month intervals. Ten percent of participants had a personal history of breast cancer and 30 percent had a family history of breast cancer; 0.1 percent were known BRCA 1/2 mutation carriers. Women with large breasts (>7cm compressed breast thickness) were excluded due to “decreased reliability” of ABUS in this group. In the overall study population, 57 breast cancers were diagnosed, 23 of which were identified by screening mammography, and 23 of which were only identified by ABUS for a cancer detection rate of 3.6 per 1,000 exams (95% CI, 2.3 to 5.4). Eleven were interval cancers that were not identified by either modality. Twenty-two of 23 breast cancers identified solely by ABUS were invasive. The stage distribution of cancers detected by ABUS in the entire group was Stage I: 17 (77.2%), Stage II: 4 (18.1%), Stage III: 1 (4.5%). Of the 23 breast cancers identified by mammography, 17 were invasive. The stage distribution of cancers identified by mammography prior to ABUS was: Stage I: 13 (76.4%), Stage II: 4 (23.5%), Stage III: 1 (5.9%). Recall rates for additional imaging were 8.7 percent (95% CI, 8.0 to 9.4) for ABUS and 4.2 percent for mammography. Biopsy rates were 1.2 percent for ABUS and 0.9 percent for mammography. An important limitation of the study was that the lead author on the study had a majority financial interest in the ABUS technology being studied, and the co-author was also a shareholder.

A fair quality multi-site study reported findings on 15, 318 women with heterogeneous or extremely dense breasts underwent concurrent mammography and ABUS.⁷⁴ Mean age was 53.3 years; 3.6 percent had a personal history of breast cancer, and 21.7 percent had a first degree relative with breast cancer. One percent were known to have a BRCA1 or BRCA2 mutation. Thirty breast cancers (28 were invasive) were detected by ABUS only for a cancer detection rate of 1.9 per 1000 exams (95% CI, 1.3 to 2.8). The recall rate was 13.5% (95% CI, 12.9 to 14.0 percent), and the biopsy rate was 3.6 percent (95% CI, 3.4 to 4.0). This study also had the important limitation that it was funded by the manufacturer of the ABUS system; six authors, including the lead author, were paid consultants for the manufacturer and three authors were shareholders in the company.

MRI

From a good-quality, multi-center U.S. prospective cohort study, we obtained unpublished data on proximate breast cancer outcomes among 334 asymptomatic women with elevated breast cancer risk and heterogeneously or extremely dense breasts in at least one mammographic quadrant (without BRCA 1/2 mutations, personal history of breast cancer, or history of chest irradiation) who underwent a single round of MRI screening.⁶¹ Elevated breast cancer risk in this subgroup was attributable mainly to family history of breast cancer. Prior to MRI screening, all women had three rounds of mammographic and breast ultrasound screening with recent negative screening mammography and ultrasound. In these 334 women, breast cancer was diagnosed by MRI in 7, including 6 invasive breast cancers and 1 DCIS (breast cancer detection rate, 21.0 cancers per 1,000 exams (95% CI, 8.5 to 42.7). In the overall sample of 8 invasive cancers detected by MRI (including women outside the dense breast subgroup), all of the invasive cancers with nodal staging were node-negative. In the subgroup with dense breasts, there were no interval breast cancers during a one year post-MRI followup period. Biopsy rates were not reported to us for the subgroup, but in the overall study sample, 7.0 percent of women screened with MRI were recommended for biopsy on account abnormal MRI findings.

Authors of a good-quality German single-center prospective cohort study provided data on a subgroup of 105 women with dense breasts undergoing MRI screening on account of dense breasts.²³ Prior to MRI screening, all women had had negative screening mammography and HHUS. The reference standard included biopsy results for positive screens and two-year prospective followup; there was no loss to followup. Among these 105 women, MRI screening detected 2 invasive breast cancers and 1 DCIS, (breast cancer detection rate 28.6 per 1,000 exams [95% CI, 5.9 to 81.2]). Both invasive cancers were Stage I. There were no interval breast cancers during the two-year followup period. Biopsy rates were not reported for the overall study sample or the subgroup.

In the good-quality, multi-center Dutch MRISC prospective cohort study, we abstracted proximate breast cancer outcomes among 1,723 elevated- risk women with BI-RADS c or d breast density without BRCA 1/2 gene mutations who underwent supplemental MRI screening along with regular screening mammography and clinical breast examination.⁶⁸ Among these women, breast cancer was detected by MRI in 6 women, for a cancer detection rate of 3.5 per 1,000 exams (95% CI, 1.3 to 7.6). Cancer stage was not reported for the dense breast subgroup, but in the overall MRISC study, MRI-detected breast cancers were more likely to be localized

than among women with incident breast cancer in the general Dutch population.⁷⁹ Overall 11.5 percent (95% CI, 10.0 to 13.1) of women with dense breasts required additional investigation or biopsy on the basis of abnormal MRI results.

It is notable that the two of three studies on supplemental MRI screening in women with dense breasts targeted women with elevated lifetime breast cancer risk.^{61, 68} Thus, the subgroup of women with dense breasts in these studies likely has elevated breast cancer risk based on risk factors other than breast density (e.g., family history of breast cancer). Only one study included a subgroup that received supplemental MRI screening solely on account of increased breast density.²³ The incidence and prevalence of breast cancer could be lower in the general population of women with dense breasts than the women represented in these studies.

DBT

Three fair-quality observational studies from the United States included data on a subgroup of women with BI-RADS c or d density.⁷⁵⁻⁷⁷ These retrospective descriptive studies compared women who had screening digital mammography with those undergoing digital mammography with concurrent DBT. Some outcomes were reported separately for exams of women with dense breasts. One study compared 4,666 exams of women with dense breasts undergoing digital mammography with DBT to a comparison group of 7,009 exams of women with dense breasts undergoing mammography alone.⁷⁷ Cancer detection rates were 5.4 per 1,000 exams (95% CI, 3.5 to 7.9) for DBT with mammography and 4.0 per 1,000 exams (95% CI, 2.7 to 5.8) in the comparison group. McCarthy et al compared outcomes of 5,056 exams of women with dense breasts undergoing DBT and mammography with a cancer detection rate of 6.9 per 1000 exams to 3,489 exams of women with dense breasts examined with mammography only with a cancer detection rates of 5.2 per 1,000 exams (95% CI, 3.1 to 8.1).⁷⁵ Invasive cancer detection rates were reported separately for the subgroup of women with dense breasts in this study: 3.4 per 1,000 exams (95% CI, 1.8 to 6.0) for digital mammography and 4.7 per 1,000 exams (95% CI, 3.0 to 7.0) for digital mammography plus DBT.⁷⁵ Recall rates were reported in all three studies and were generally lower in women examined with both DBT and mammography. Recall rates were 9.1% (95% CI, 8.4 to 11.0) for digital mammography only compared with 6.9 percent (95% CI, 6.2 to 7.7) with DBT and mammography⁷⁷; 12.8 percent (95% CI, 11.7 to 14.0) for mammography only compared with 10.8 percent (95% CI, 10.0 to 11.7) for DBT with mammography⁷⁵; and 16.6 percent (95% CI, 15.0 to 18.2) for mammography only compared with 9.7 percent (95% CI, 8.6 to 11.0) with DBT combined with digital mammography.⁷⁷ Biopsy rates were not reported separately for the subgroup of women with dense breasts, but biopsy rates for the overall group, when reported, were similar for women examined with or without DBT.^{75, 77}

A fair-quality study from Italy compared the results of digital mammography screening with the results of mammography with DBT in a prospective cohort of 7,294 exams of 7,292 women (median age 58 years); 1,215 exams were done on women with dense breasts.⁷⁰ All exams were read independently by two radiologists. Each read the digital mammogram first and subsequently interpreted mammogram together with the DBT images. Women were recalled for further evaluation if radiologists recommended recall at either mammography or at DBT. In the subgroup of women with dense breasts, cancer detection rates were 4.1 per 1,000 exams (95%

CI, 1.3 to 9.6) based on mammography interpretation and 6.6 (95% CI, 2.9 to 12.9) per 1,000 exams based on mammography plus DBT. Recall rates in the subgroup of women with dense breasts were 7.2 percent (95% CI, 5.8 to 8.8) based on digital mammography and declined to 6.6 percent (95% CI, 5.3 to 8.1) with review of DBT images. Biopsy rates were not reported separately.

Key Question 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?

Summary of Results

In 24 states, mammography providers are required by law to notify women if they have dense breasts.³⁷ Evidence on the harms of notification of breast density and the harms of supplemental testing for women with dense breasts was sparse.

A single good-quality RCT from Canada examined the effects of notification to women of mammographic findings of dense breasts (defined as greater than 50 percent of breast volume).⁸⁰ In this study, 285 women randomized to the intervention group received a report of their breast density as part of the letters summarizing their mammography results, as well as pamphlets on breast cancer risk factors (including density). No supplemental screening was recommended. At 4 weeks, more women in the intervention group had statistically significantly increased knowledge of breast density, more perceived themselves to have an elevated breast cancer risk, and more stated that they were very likely to have a clinical breast exam. These differences did not persist at 6 months. There were no differences in psychological distress, breast cancer worry, or preoccupation with breast cancer at 4 weeks or 6 months.

In addition to detecting more cancers, supplemental screening with HHUS, ABUS, and MRI leads to more recalls and biopsies for false positive results, as summarized in KQ's 2 and 3 (**Tables 13-19**). In women with negative screening mammography, the positive predictive value of HHUS ranged from 3.0 to 8.3 percent, meaning over 90 percent of positive tests were false positive. In good-quality studies, biopsy rates associated with supplemental screening tests ranged from 5.9 to 6.7 percent (**Table 16**). For ABUS, the PPV in one study was 4.1 percent, and in two studies of ABUS, recall rates were 1.5 percent and 8.7 percent (**Table 17**). Positive predictive value for three good-quality studies of breast MRI ranged from 3.0 to 33.3 percent, but women in the MRI screening studies were generally selected for elevated breast cancer risk based on family history apart from increased breast density. In these studies, recall rates ranged from 8.6 to 23.4 percent (**Table 18**).

DBT led to reduced recall rates for women with dense breasts in four fair-quality studies (**Table 19**), but when added to digital mammography more than doubles the radiation exposure from each screening exam.^{28, 29, 31} Technology that allows reconstruction of the 2-D breast images can

reduce radiation exposure but is not widely disseminated.³¹

The evidence on harms of recall, biopsy, and overdiagnosis after mammography screening are summarized in *Screening for Breast Cancer: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation*.⁸¹ Although we found no studies specifically addressing the harms related to false positives and overdiagnosis for supplemental screening for women with dense breasts, they are likely to be very similar to those noted for mammography, with the exception of harms from increased radiation exposure with DBT and gadolinium contrast used in MRI screening.

We found no specific reports of adverse effects of breast MRI, but it does require the administration of gadolinium contrast. Gadolinium contrast media have been associated with nephrogenic systemic fibrosis in patients with acute kidney injury or chronic kidney disease (CKD) Stage 4 and 5, and gadolinium is considered contraindicated in these patients. Nephrogenic systemic fibrosis is characterized by scleroderma-like tissue changes in the skin, internal organs, eyes, and blood vessels,⁸² is associated with increased mortality and has no effective treatment.⁸³ The American College of Radiology recommends screening with serum creatinine prior to administration of gadolinium for those with meeting any of the following criteria: age over 60 years, history of hypertension, medication use for diabetes mellitus, known history of kidney disease or surgery.⁸⁴ The need for additional screening for CKD to reduce potential risk of nephrogenic systemic fibrosis is part of the consideration of potential harms related to breast MRI, particularly for older women and women with hypertension and/or diabetes.

Given the sparse literature on harms on notification of dense breasts, and the mandates for notification in nearly half of all states, additional studies of the behavioral impacts of breast density notification in the United States are needed. Longer term studies are also needed on outcomes of supplemental screening, including recall and biopsy rates of recurrent screening, overdiagnosis, and frequency of harms associated with gadolinium contrast administration for breast MRI.

Study Details

In a good-quality study of possible psychological harms and other effects of breast density notification, the screening mammography program of British Columbia conducted an RCT of notification of breast density findings at the time of mailed notification of screening mammography results (**Table 22**).⁸⁰ Women were recruited at the time of appointment scheduling for mammography screening; women with suspicious mammography findings or breast density exceeding 50 percent at the previous exam were excluded from the study. The final sample included 618 women, 333 in the control group and 285 in the intervention group, and provided >90 percent power with $\alpha < .05$ in the *a priori* sample size calculation. The groups were similar on baseline measures. Women in the intervention group and their primary physicians received information about their breast density with mailed results of their mammogram, along with an educational pamphlet. Supplemental screening was not recommended mentioned in the pamphlet, as it was not routinely available to women in the screening program. Women in the control group and their primary physicians received the usual

results letter. Prior to the screening mammogram and at 4 weeks and 6 months following screening, women in both groups were interviewed by telephone about anxiety, depression, breast cancer worry, self-estimates of breast cancer risk, and knowledge of breast density as a risk factor. Women were also queried about whether they had a recent clinical breast exam or had plans to have one, dietary changes, and their intention to return for mammography at the recommended time interval.

Survey completions rates at 4 weeks and 6 months were 93.7 percent and 94.8 percent with no differential loss to followup. No differences were found between groups in psychological measures (the primary study outcome) at 4 weeks or 6 months (**Table 23**). At 4 weeks, fewer women in the intervention group (10.5%) compared with the control group (15.5%) perceived themselves to be at “a lot lower risk” for breast cancer relative to other women. This difference was statistically significant at 4 weeks, but not at 6 months. At 4 weeks, 24.8 percent of women in the intervention group correctly described breast density, compared with 7.5 percent in the control group ($p<.001$), and more women in the intervention group knew that it was a risk factor for breast cancer (85.3% vs. 66.4%; $p<.001$). Twenty-three percent of the intervention group compared with 15.1 percent of the control group indicated intention to have a clinical breast exam at the 4 week interview ($p=.03$), but at 6 months this difference was not sustained.

Chapter 4. Discussion

Summary of Evidence

We conducted this systematic review to assist the USPSTF in consideration of a recommendation regarding supplemental breast cancer screening after negative mammography for women found to have dense breasts (**Table 24**). While there are other important risk factors for breast cancer, some with higher relative risks, breast density is unique in that many states have mandated direct-to-consumer communication of mammographic breast density findings. This raises questions for women and their doctors about the interpretation of screening results and the need for additional testing. This review examined the accuracy and reproducibility of BI-RADS determination of breast density. Evidence was also reviewed on test performance characteristics and breast cancer detection rates, interval cancer rates, and rates of recall and biopsy for specific supplemental screening modalities (hand-held ultrasound [HHUS], automated whole breast ultrasound [ABUS], magnetic resonance imaging [MRI], and digital breast tomosynthesis [DBT]). The sparse evidence on the harms associated with breast density notification was summarized.

BI-RADS Breast Density Determinations

We examined the reproducibility of categorical BI-RADS breast density determinations in U.S. clinical-community practices since this is the system recommended by American College of Radiology and is written into most of the legislative mandates. The reproducibility of sequential categorizations of breast density by the same radiologist was better than that for different readers, but still lower than desirable. Best estimates from these data suggest about one in five women would be categorized into a different BI-RADS density category (a, b, c, d) by the same radiologist after the second screening exam, while one in three would be categorized differently after a second exam read by a different radiologist. Re-interpretation of breast density findings can occur due to multiple factors stemming from the woman being examined, the qualitative nature of the technique, and the expertise of the radiologist interpreting the exams. While radiologists' breast density readings are an important part of the screening examination, the impact of variability in exam readings may have unintended consequences through direct-to-consumer communications. Reclassifications from one category to another may undermine women's confidence in the screening process and leave them uncertain about their risk for breast cancer, but would not necessarily result in changes in recommended clinical actions or the content of mandated communications depending on individual state mandates. The American College of Radiology has publically expressed similar concerns.⁸⁵ A smaller but still sizeable proportion of women (about 13 to 19 percent, depending on whether the second exam was read by the same or a different radiologist) would be reclassified into a different overall combined category (i.e., from "non-dense" to "dense" or vice versa). In these instances, different mandated communications about elevated breast cancer risk and/or the need for additional clinical screenings could occur for the same woman in the span of 2 to 3 years. Finally, estimates based on overall programmatic-level reclassifications could be magnified in certain circumstances. Test-retest reproducibility among individual community-based radiologists suggests the potential

for even greater reclassification impacts, particularly given outlier performers.. For example, in one test set read by U.S.-based radiologists, of the 145 mammograms read as “non-dense” (a or b) by the majority of readers, 83 percent were interpreted as dense by at least one radiologist. Similarly, almost half (47%) of the 187 mammograms read as “dense” by the majority were read as “non-dense” by at least one radiologist. In fact, almost all (93%) of the mammograms were interpreted as dense (c or d) by at least one radiologist.⁵⁷

Valid and reliable breast density calculations could also be important in the development of more personalized breast cancer screening recommendations. A recent model examining average-risk as well as higher-risk women (i.e., those with family or personal history of breast cancer or previous breast biopsy) found optimal screening strategies vary based on BI-RADS density determinations as well as age.⁸⁶ These modeled results suggest that, depending on the selected threshold for decision-making and the woman’s age, BI-RADS category, and history, optimal mammography screening would start at age 40 or 50, with repeat screenings every 2 to 3 or 4 years, and reassessment at each decade through age 70 years. Clearly, valid and reproducible BI-RADs breast density classification would be important for implementation of such personalized screening approaches.

Our findings are consistent with cautions from the American College of Radiology about benefits, possible harms, and unintended consequences for the communication of breast density assessments to women.⁸⁵ While some of the reclassifications may appropriately reflect actual physiological changes in breast density over time, it is unlikely that the majority of findings reflect true breast density differences due to design of the studies and the consistency of findings. The community practice data may also be criticized for representing relatively “older” practice (i.e., radiologist readings primarily from 2000 to 2009). However, concerns about the reproducibility of BI-RADS breast density determinations are not new, and have been a major impetus for research examining other more objective methods for assigning breast density such as ultrasonographic assessments, automated volumetric estimations, or other computer-assisted methods. Although variability is reduced by use of double readings, which is widely practiced in Europe,⁷⁸ this approach may not be practical in the United States due to 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.

Supplemental Test Performance and Outcomes for Women With Dense Breasts and a Negative Mammogram

There was limited evidence on the test performance characteristics of potential supplemental screening modalities for women identified as having dense breasts. Two good-quality studies of HHUS were relatively consistent in estimates of sensitivity, specificity, and positive predictive value. Smaller good-quality studies of high risk women suggest that the sensitivity of MRI for women with dense breasts is likely somewhat higher than HHUS. Specificity of these modalities is similar, and PPV is low, resulting in many false positive recalls and biopsies (**Table 21**). No summary is possible for the other modalities, with only a single study of ABUS and no studies of DBT. We found no studies of any supplemental test that examined the effect of age and other breast cancer risk factors on test performance characteristics in women with dense breasts.

Evidence on important clinical outcomes of supplemental screening among women with dense breasts was also limited. No studies examined the most important outcomes: breast cancer recurrence rates and mortality rates with and without supplemental screening. In general, supplemental screening consistently found additional breast cancers not identified by mammography. The substantial majority of breast cancers identified by MRI and HHUS were invasive. Invasive cancer detection rates were lower in the two good-quality studies of HHUS (3.9 and 4.1 per 1,000 exams)^{61, 62} than in three good-quality studies of breast MRI (3.4 to 12.3 per 1,000 exams)^{23, 61, 68} (**Table 21**) but this may reflect a higher baseline risk in the women recruited to MRI studies. Recall rates and biopsy rates were markedly increased by supplemental screening with HHUS and MRI. Limited evidence on DBT suggested that recall rates for additional imaging are reduced in women with dense breasts; however, it is unclear if this results in fewer overall breast biopsies. Evidence on ABUS in women with dense breasts was limited; both U.S. studies were industry-funded.

Harms of Breast Density Notification and Supplemental Testing

Evidence on harms of breast density notification and harms resulting from supplemental screening were very limited. Only one trial was identified which suggested little behavioral or psychological impact of breast density notification at six months of followup. Additional studies of the effects of breast density notification are needed.

No studies of supplemental screening addressed the important potential risks of overdiagnosis with the associated harms of unnecessary treatment. Additional studies in U.S. populations are needed. False positive tests, and their associated harms, including anxiety and unnecessary biopsies are increased by supplemental HHSU and MRI; evidence on ABUS and DBT is too limited to draw any conclusions.

Limitations of the Review

Limitations Due to Our Approach

We reviewed only studies published in English. It is possible that studies published in other languages would have met our inclusion criteria, although applicability to U.S. practice may have been reduced, particularly for BI-RADS density reproducibility studies. For applicability as well as feasibility concerns, we did not review other approaches to breast density assessment. We are aware of additional studies of supplemental screening of women with dense breasts with MRI or ABUS that are in process (**Appendix B**), but we were limited to the published literature.

Limitations Due to the Evidence Base

Current evidence suggests that the use of BI-RADS categorization to classify breast density in women results in clinically important reclassifications with repeat imaging. This raises questions about whether women are being appropriately selected for supplemental screening. The number, quality, and rigor of studies of test characteristics and clinical outcomes were quite limited. We

identified no long term followup comparative studies of the clinical outcomes or harms of supplemental screening for women with dense breasts. Many studies did not apply a complete reference standard or a clear description of followup, so we were unable to calculate test performance characteristics. Only one comparative study of cohorts with and without supplemental screening adjusted for differences between cohorts. Most studies assessed short-term incremental impact among women undergoing screening mammography and supplemental screening. Many studies included women at increased breast cancer risk due to risk factors other than breast density, limiting the generalizability of findings in these studies to the general screening population of women with dense breasts. Studies of breast MRI focused on women with multiple risk factors; very little data were available for women with dense breasts and no other major risk factors for breast cancer. Literature on ABUS and breast tomosynthesis for women with dense breasts was very limited and more studies are needed before any conclusions can be reached. Literature on harms related to identification of women as having dense breasts or other potential harms related to supplemental screening was very sparse.

Future Research Needs

BI-RADS breast density categorization is a subjective assessment and no clear objective reference standard for breast density has been established. The effect of double-reading with consensus and other quality measures on reproducibility of readings should be studied. Automated computer assessments of BI-RADS density have been developed but none have come into widespread clinical use in the United States. A research base to establish clinically applicable approaches, including automated systems, would require both internal consistency and reliability, predictive association with breast cancer risk, as well as reasonable correlation with the BI-RADS density assessments of experienced breast radiologists.

Longer term studies with comprehensive followup, preferably randomized trials, are needed to understand the impact of supplemental testing in women with dense breasts on important breast cancer outcomes including morbidity and mortality. Only with RCTs and longer term followup can the risks of length time bias (earlier detection of cancer not resulting in improved outcomes) as well as the impact of overdiagnosis (leading to unnecessary treatment) be evaluated.

Such studies could evaluate the impact of supplemental screening for women with dense breasts on breast cancer stage at diagnosis, recurrence rates, mortality and overdiagnosis rates. Studies are especially needed among general populations of women with dense breasts who are not otherwise selected for elevated breast cancer risk. Rigorous documentation of the effects of supplemental testing on interval breast cancer rates, recurrence rates, and recall and biopsy rates with repeated screening would assist in elucidating the benefits and harms of supplemental screening for women with dense breasts.

Conclusion

Good-quality studies with U.S. radiologists show important reclassification between dense and non-dense breasts in at least 12.6 to 18.7 percent of women undergoing sequential screening

examinations. Reclassification of breast density may introduce confusion or reduce confidence among women receiving mandated breast density notifications. Moving from a “dense” to a “non-dense” breast categorization will result in different mandated communications in many states with breast density legislation as well as differences in recommended or considered supplemental screening approaches. As risk-based personalized screening approaches develop, valid and reproducible breast density categorization will be required.

Existing limited evidence suggests that more breast cancers will be detected by supplemental HHUS and MRI screening of women with dense breasts, and that most detected breast cancers will be invasive. Whether diagnosis of additional breast cancers leads to improved clinical outcomes, or what proportion of the cancers diagnosed represent overdiagnosis has not been evaluated. However, supplemental testing of women with dense breasts with HHUS or MRI is clearly associated with a marked increase in recall rates for diagnostic investigation among women who do not have breast cancer. DBT may not increase recall rates, but evidence is too limited for definitive conclusions. There is a critical need for well-designed, long-term comparative studies of supplemental screening of women with dense breasts, so that meaningful clinical outcomes of supplemental screening of women with dense breasts can be characterized.

References

1. Wolfe JN. Breast patterns as an index of risk for developing breast cancer. *AJR Am J Roentgenol.* 1976;126(6):1130-7. PMID: 179369
2. Byng JW, Boyd NF, Fishell E, Jong RA, Yaffe MJ. The quantitative analysis of mammographic densities. *Phys Med Biol.* 1994;39(10):1629-38. PMID: 15551535
3. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, et al. The ACR BI-RADS experience: learning from history. *J Am Coll Radiol.* 2009;6(12):851-60. PMID: 19945040
4. American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS). Reston, VA: American College of Radiology; 1992.
5. Sickles E, D'Orsi C, Bassett LW, et al. ACR BI-RADS® Mammography. In: ACR BI-RADS® Atlas, Breast Imaging and Reporting Data System Reston, VA: American College of Radiology 2013.
6. Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, et al. Mammographic density, breast cancer risk and risk prediction. *Breast Cancer Res.* 2007;9(6):217. PMID: 18190724
7. 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(10). PMID: 25217577
8. Kerlikowske K, Zhu W, Tosteson AN, Sprague BL, Tice JA, Lehman CD, et al. Identifying women with dense breasts at high risk for interval cancer: a cohort study. *Ann Intern Med.* 2015;162(10):673-81. PMID: 25984843
9. Mousa A, Ryan E, Mello-Thoms C, Brennan P. What effect does mammographic breast density have on lesion detection in digital mammography? *Clin Radiol.* 2014;69(4):333-41. PMID: 24424328
10. Boyd NF, Byng JW, Jong RA, Fishell EK, Little LE, Miller AB, et al. Quantitative Classification of Mammographic Densities and Breast Cancer Risk: Results From the Canadian National Breast Screening Study. *J Natl Cancer Inst.* 1995;87(9):670-5. PMID: 7752271
11. Boyd NF, Lockwood GA, Byng JW, Trichler DL, Yaffe MJ. Mammographic densities and breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 1998;7(12):1133-44. PMID: 9865433
12. Byrne C, Schairer C, Wolfe J, Parekh N, Salane M, Brinton LA, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. *J Natl Cancer Inst.* 1995;87(21):1622-9. PMID: 7563205
13. McCormack VA, dos Santos Silva I. Breast Density and Parenchymal Patterns as Markers of Breast Cancer Risk: A Meta-analysis. *Cancer Epidemiol Biomarkers Prev.* 2006;15(6):1159-69. PMID: 16775176
14. Price ER, Hargreaves J, Lipson JA, Sickles EA, Brenner RJ, Lindfors KK, et al. The California breast density information group: a collaborative response to the issues of breast density, breast cancer risk, and breast density notification legislation. *Radiology.* 2013;269(3):887-92. PMID: 24023072

15. Pettersson A, Graff RE, Ursin G, Santos Silva ID, McCormack V, Baglietto L, et al. Mammographic density phenotypes and risk of breast cancer: a meta-analysis. *J Natl Cancer Inst.* 2014;106(5). PMID: 24816206
16. 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(16):1218-27. PMID: 22911616
17. Carmona-Sanchez E, Cuadros Lopez JL, Cuadros Celorrio AM, Perez-Roncero G, Gonzalez Ramirez AR, Fernandez Alonso AM. Assessment of mammographic density in postmenopausal women during long term hormone replacement therapy. *Gynecol Endocrinol.* 2013;29(12):1067-70. PMID: 24004297
18. Hou N, Hong S, Wang W, Olopade OI, Dignam JJ, Huo D. Hormone replacement therapy and breast cancer: heterogeneous risks by race, weight, and breast density. *J Natl Cancer Inst.* 2013;105(18):1365-72. PMID: 24003037
19. Tice JA, Cummings SR, Smith-Bindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. *Ann Intern Med.* 2008;148(5):337-47. PMID: 18316752
20. Del Carmen MG, Hughes KS, Halpern E, Rafferty E, Kopans D, Parisky YR, et al. Racial differences in mammographic breast density. *Cancer.* 2003;98(3):590-6. PMID: 17377060
21. McCormack VA, Perry N, Vinnicombe SJ, Silva Idos S. Ethnic variations in mammographic density: a British multiethnic longitudinal study. *Am J Epidemiol.* 2008;168(4):412-21. PMID: 18621673
22. Centers for Disease Control and Prevention. Breast Cancer Rates by Race and Ethnicity 2014 [March 3, 2015]; Available from: <http://www.cdc.gov/cancer/breast/statistics/race.htm>.
23. 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(22):2304-10. PMID: 24958821
24. Saslow D, Boetes C, Burke W, Harms S, Leach MO, Lehman CD, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer Journal for Clinicians.* 2007;57(2):75-89. PMID: 17392385
25. Mainiero MB, Lourenco A, Mahoney MC, Newell MS, Bailey L, Barke LD, et al. ACR Appropriateness Criteria Breast Cancer Screening. *J Am Coll Radiol.* 2013;10(1):11-4. PMID: 23290667
26. Lee SH, Im RJ, Yoon CS, Hur MH, Lee HK, Kang SS, et al. Initial experience of breast specific gamma imaging (BSGI) in Asia. *European Journal of Cancer, Supplement.* 2010;8(3):238
27. 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(2):166-71. PMID: 22098941
28. Feng SS, Sechopoulos I. Clinical digital breast tomosynthesis system: dosimetric characterization. *Radiology.* 2012;263(1):35-42. PMID: 22332070
29. Olgar T, Kahn T, Gosch D. Average glandular dose in digital mammography and breast tomosynthesis. *Rofo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin.* 2012;184(10):911-8. PMID: 22711250

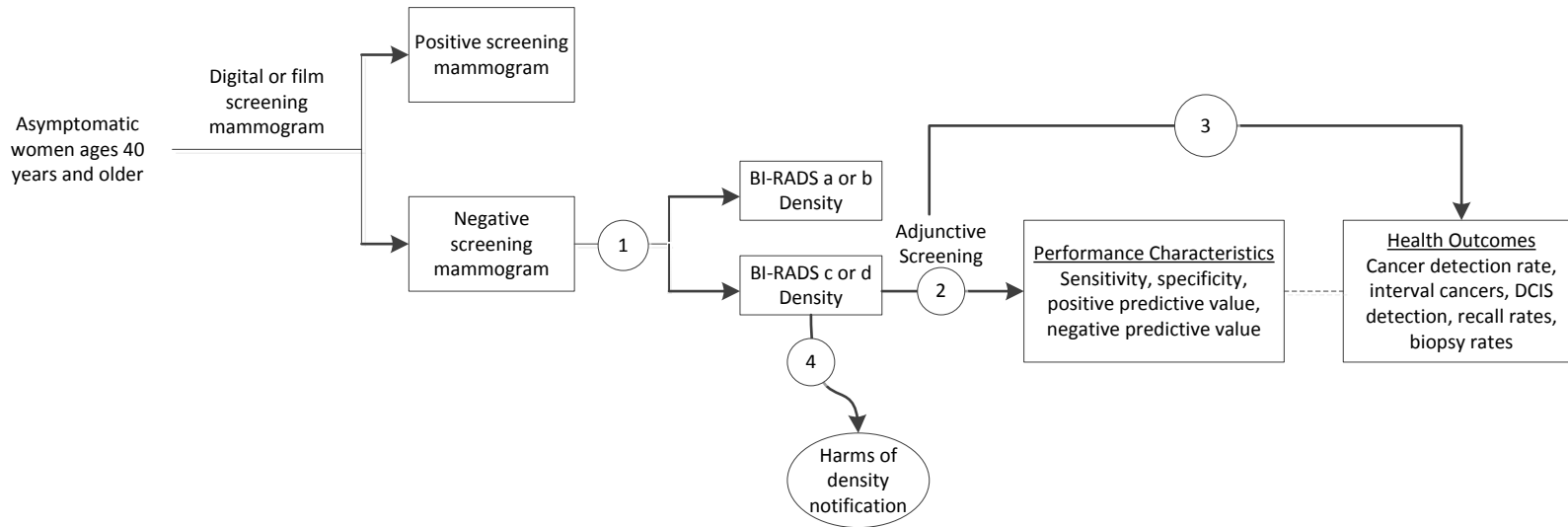
30. Skaane P, Bandos AI, Gullien R, Eben EB, Ekseth U, Haakenaasen U, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology*. 2013;267(1):47-56. PMID: 23297332
31. Gur D, Zuley ML, Anello MI, Rathfon GY, Chough DM, 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(2):166-71. PMID: 22098941
32. Zuley ML, Guo B, Catullo VJ, Chough DM, Kelly AE, Lu AH, et al. Comparison of Two-dimensional synthesized Mammograms versus Original Digital Mammograms alone and in combination with Tomosynthesis images. *Radiology*. 2014;271(3):664-71. PMID: 24475859
33. Skaane P, Bandos AI, Eben EB, Jepsen IN, Krager M, Haakenaasen U, et al. Two-View Digital Breast Tomosynthesis Screening with Synthetically Reconstructed Projection Images: Comparison with Digital Breast Tomosynthesis with Full-Field Digital Mammographic Images. *Radiology*. 2014;271(3):655-63. PMID: 24484063
34. National Comprehensive Cancer Network (NCCN). Breast Cancer Screening and Diagnosis. 2014 [Accessed October 10, 2014]; Available from: http://www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf.
35. American Cancer Society (ACS). American Cancer Society Recommendations for Early Breast Cancer Detection in Women Without Breast Symptoms. 2014 [Accessed October 7, 2014]; Available from: <http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-ac-s-recs>.
36. American College of Obstetricians and Gynecologists. Committee Opinion No. 625: Management of Women With Dense Breasts Diagnosed by Mammography. *Obstet Gynecol*. 2015;125(3):750-1.25730253
37. Are you Dense? Advocacy. D.E.N.S.E. ® State Efforts. 2014 [Accessed October 7, 2014]; Available from: <http://www.areyoudenseadvocacy.org/dense/>.
38. 77th Oregon Legislative Assembly. SB 420 2013 [Accessed January 2015]; Available from: <https://olis.leg.state.or.us/liz/2013R1/Downloads/MeasureDocument/SB420/Enrolled>.
39. Illinois General Assembly. (215 ILCS 5/356g) Sec. 356g. Mammograms; mastectomies. [October 10, 2014]; Available from: <http://www.ilga.gov/legislation/ilcs/documents/021500050k356g.htm>.
40. The State of Connecticut. Public Act No. 09.41. An Act Requiring Communication of Mammographic Breast Density Information to Patients. 2009 [Accessed October 10, 2014]; Available from: <http://www.cga.ct.gov/2009/ACT/PA/2009PA-00041-R00SB-00458-PA.htm>.
41. 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 October 10, 2014]; Available from: http://www.njleg.state.nj.us/2012/Bills/PL13/196_PDF.
42. Indiana 118th General Assembly. Senate Enrolled Act No. 414. An Act to amend the Indiana Code concerning insurance. 2013 [Accessed October 10, 2014]; Available from: <http://www.in.gov/legislative/bills/2013/PDF/SE/SE0414.1.pdf>.
43. Alkhenizan A, Hussain A, Alsayed A. The sensitivity and specificity of screening mammography in primary care setting in Saudi Arabia. *J Clin Oncol*. 2013;31(15). PMID:

44. Allen MW, Hendi P, Bassett L, Phelps ME, Gambhir SS. A study on the cost effectiveness of sestamibi scintimammography for screening women with dense breasts for breast cancer (Structured abstract). *Breast Cancer Res Treat* [serial on the Internet]. 1999 [cited Database Search; 55(3): Available from: <http://onlinelibrary.wiley.com/o/cochrane/cleed/articles/NHSEED-21999001689/frame.html>.
45. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2009;151(10):716-26, w-236. PMID: 19920272
46. U.S. Food and Drug Administration. Radiation-Emitting Products: MQSA National Statistics. 2015 [Accessed January 29, 2015]; Available from: <http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/FacilityScorecard/ucm113858.htm>.
47. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med*. 2001;20(3 Suppl):21-35. PMID: 11306229
48. U.S. Preventive Services Taskforce. U.S. Preventive Services Taskforce Procedure Manual. Rockville, MD: Agency for Healthcare Research and Quality; 2013.
49. National Institute for Health and Clinical Excellence. The Guidelines Manual. London: National Institute for Health and Clinical Excellence; 2006.
50. 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(8):854-61. PMID: 20056381
51. Clopper CJ, Pearson ES. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*. 1934;26(4):404-13
52. Brown LD, Cai TT, DasGupta A. Interval Estimation for a Binomial Proportion. *Statistical Science*. 2001;16(2):101-17
53. Melnikow J, Fenton J, DL M, Whitlock E, Weyrich M. Screening for Breast Cancer with Digital Breast Tomosynthesis Rockville, MD: Agency for Healthcare Research and Quality in press.
54. Miglioretti D, Lange J, van Ravesteyn N, van den Broek J, Lee C, Melnikow J, et al. Radiation-Induced Breast Cancer and Breast Cancer Death from Mammography Screening. Rockville, MD: Agency on Healthcare Research and Quality in press.
55. Harvey JA, Gard CC, Miglioretti DL, Yankaskas BC, Kerlikowske K, Buist DS, et al. Reported mammographic density: film-screen versus digital acquisition. *Radiology*. 2013;266(3):752-8. PMID: 23249570
56. 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(4):326-33. PMID: 22607064
57. Gard CC, Bowles EJA, Miglioretti DL, Taplin SH, CM R. Misclassification of Breast Imaging Reporting and Data System (BI-RADS) mammographic density and implications for breast density reporting legislation. *Breast J*. 2015; epub ahead of print
58. Redondo A, Comas M, Macia 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(1019):1465-70. PMID: 22993385

59. 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. *Il Giornale di chirurgia*. 2012;33(3):81-5. PMID: 22525552
60. 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 Journal*. 2012;18(4):326-33. PMID: 22607064
61. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. 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(13):1394-404. PMID: 22474203
62. 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(7):1021-6. PMID: 21211962
63. 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(1):59-69. PMID: 22723501
64. 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(10):432-9. PMID: 23138139
65. 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(4):667-75. PMID: 20853108
66. 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(3):734-42. PMID: 19727744
67. Giuliano V, Giuliano C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. *Clin Imaging* 2013;37(3):480-6. PMID: 23116728
68. 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(1):109-19. PMID: 16791481
69. 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(6):562-6. PMID: 18338490
70. Girardi V, Tonegutti M, Ciatto S, Bonetti F. Breast ultrasound in 22,131 asymptomatic women with negative mammography. *Breast*. 2013;22(5):806-9. PMID: 23558244
71. Parris T, Wakefield D, Frimmer H. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. *Breast J* 2013;19(1):64-70. PMID: 23240937
72. 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(2):347-55. PMID: 23579052
73. Weigert J, Steenbergen S. The connecticut experiment: the role of ultrasound in the screening of women with dense breasts. *Breast J*. 2012;18(6):517-22. PMID: 23009208

74. Brem RF, Tabar 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(3):663-73
75. 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(11). PMID: 25313245
76. 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(3):694-700. PMID: 23901124
77. 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 American Journal of Roentgenology*. 2013;200(6):1401-8. PMID: 23701081
78. 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. *The Lancet Oncology*. 2013;14(7):583-9. PMID: 23623721
79. Kriege M, Brekelmans CT, Boetes C, Besnard PE, Zonderland HM, Obdeijn IM, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med*. 2004;351(5):427-37. PMID: 15282350
80. Bortorff 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(1):38-57. PMID: 17450704
81. 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. Rockville, MD: Agency For Healthcare Research and Quality 2015.
82. Crownover BK, Bepko JL. Appropriate and safe use of diagnostic imaging. *Am Fam Physician*. 2013;87(7):494-501. PMID: 23547591
83. Perazella MA. Current status of gadolinium toxicity in patients with kidney disease. *Clin J Am Soc Nephrol*. 2009;4(2):461-9. PMID: 19201920
84. Besheli LD, Aran S, Shaqdan K, Kay J, Abujudeh H. Current status of nephrogenic systemic fibrosis. *Clin Radiol*. 2014;69(7):661-8. PMID: 24582176
85. American College of Radiology. ACR Statement on Reporting Breast Density in Mammography Reports and Patient Summaries April 24, 2012 [Accessed October 10, 2014]; Available from: <http://www.acr.org/About-Us/Media-Center/Position-Statements/Position-Statements-Folder/Statement-on-Reporting-Breast-Density-in-Mammography-Reports-and-Patient-Summaries>.
86. Schousboe JT, Kerlikowske K, Loh A, Cummings SR. Personalizing mammography by breast density and other risk factors for breast cancer: analysis of health benefits and cost-effectiveness. *Ann Intern Med*. 2011;155(1):10-20. PMID: 21727289

Figure 1. Analytic Framework

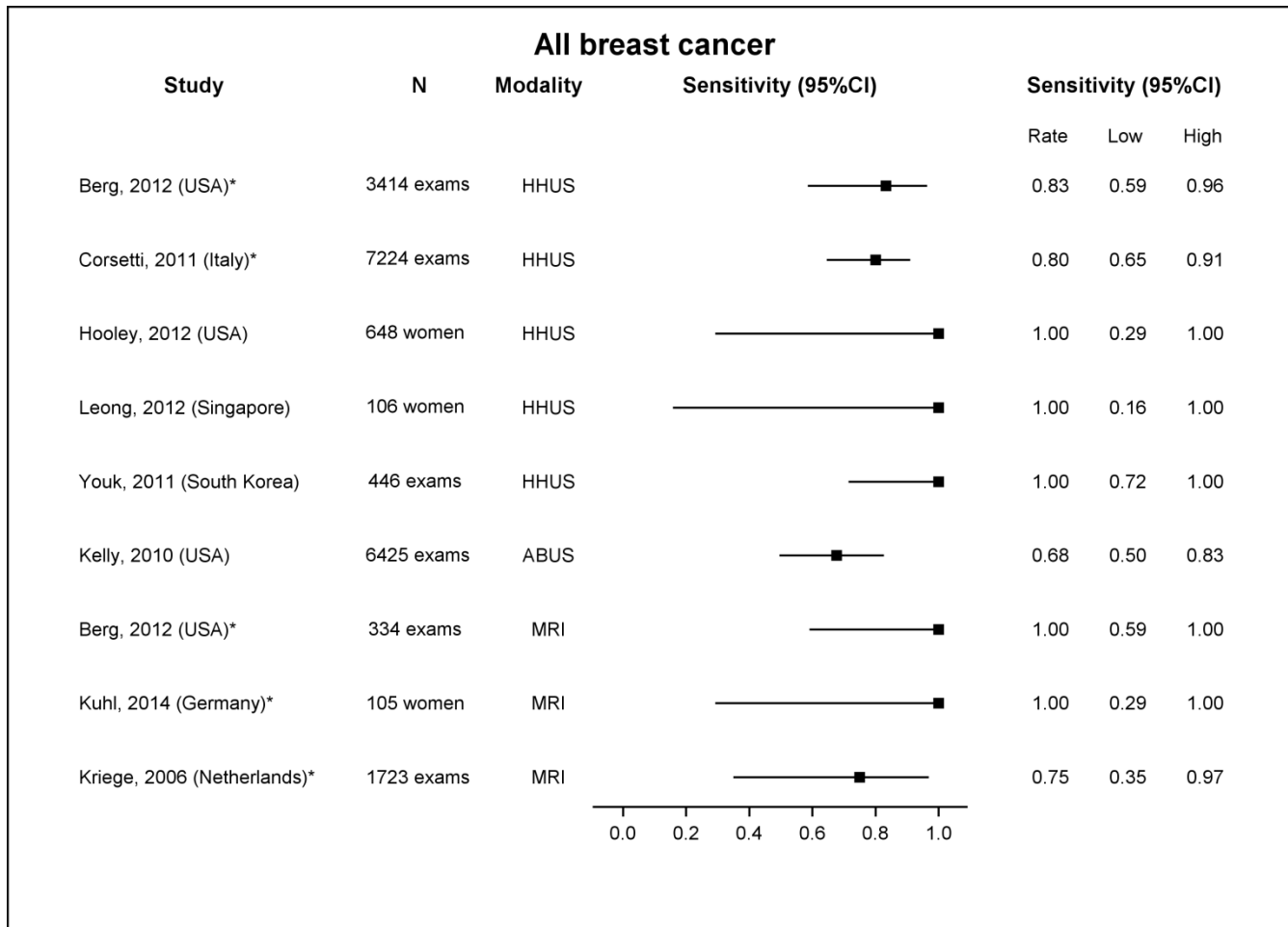


Key Questions:

1. What is the accuracy and reproducibility of BI-RADS determination of breast density?
2. What are the test performance characteristics of newer technologies for breast cancer screening when used as supplemental tests after a negative screening mammography exam in women found to have dense breasts and how do these performance characteristics differ by age and risk factors?
3. When performed after a negative screening mammogram in women found to have dense breasts, what is the effectiveness of supplemental screening with breast ultrasound, 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?
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?

Abbreviations: BI-RADS=Breast Imaging-Reporting and Data System; DCIS=ductal carcinoma in situ.

Figure 2. Sensitivity of Supplemental HHUS, ABUS, and MRI in Detecting All Breast Cancer

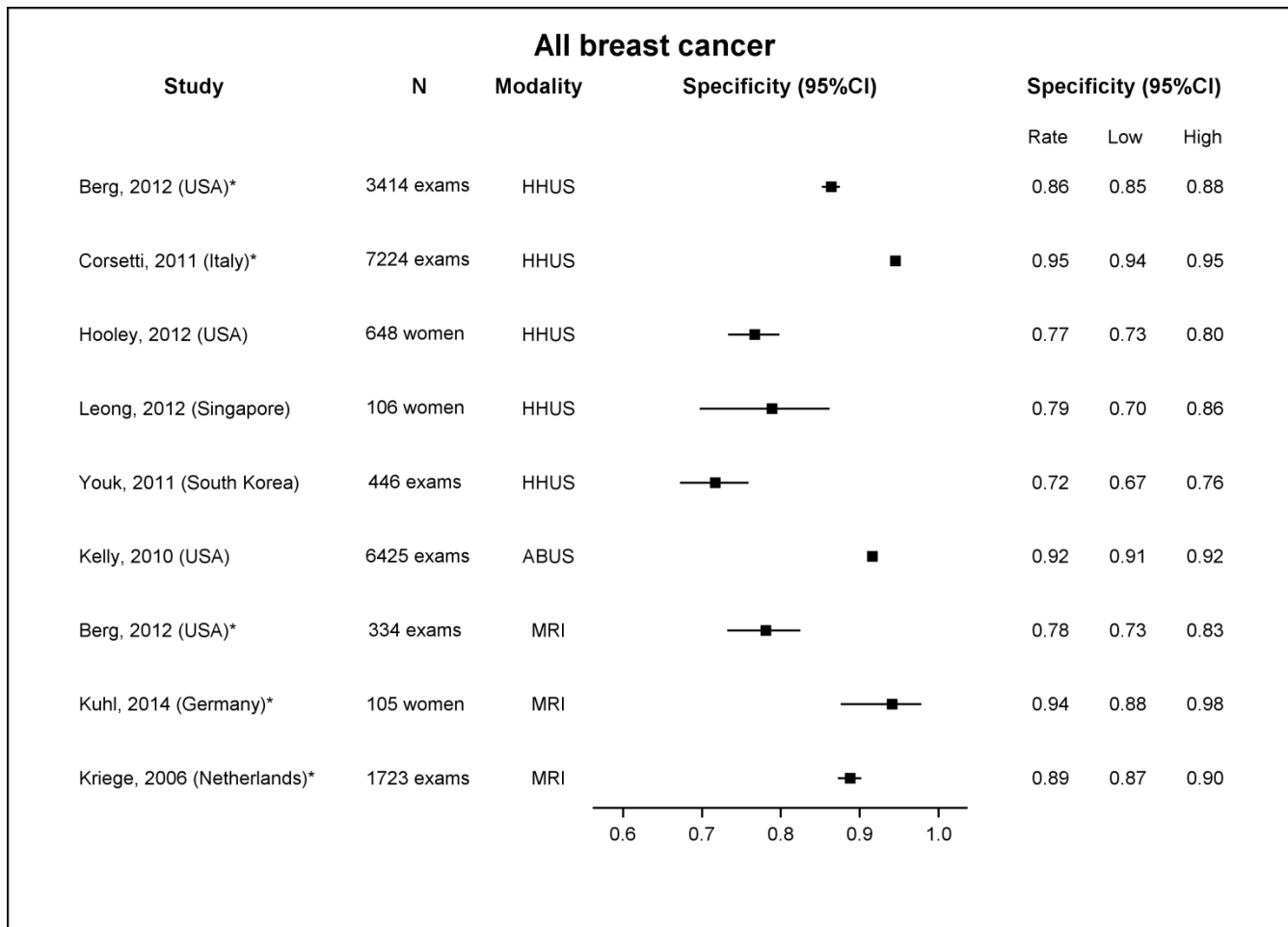


Note: DCIS cases are included as “All Breast Cancer.”

*Good-quality study .

Abbreviations: ABUS=automated whole breast ultrasound; HHUS=hand-held ultrasound.

Figure 3. Specificity of Supplemental HHUS, ABUS, and MRI in Detecting All Breast Cancer

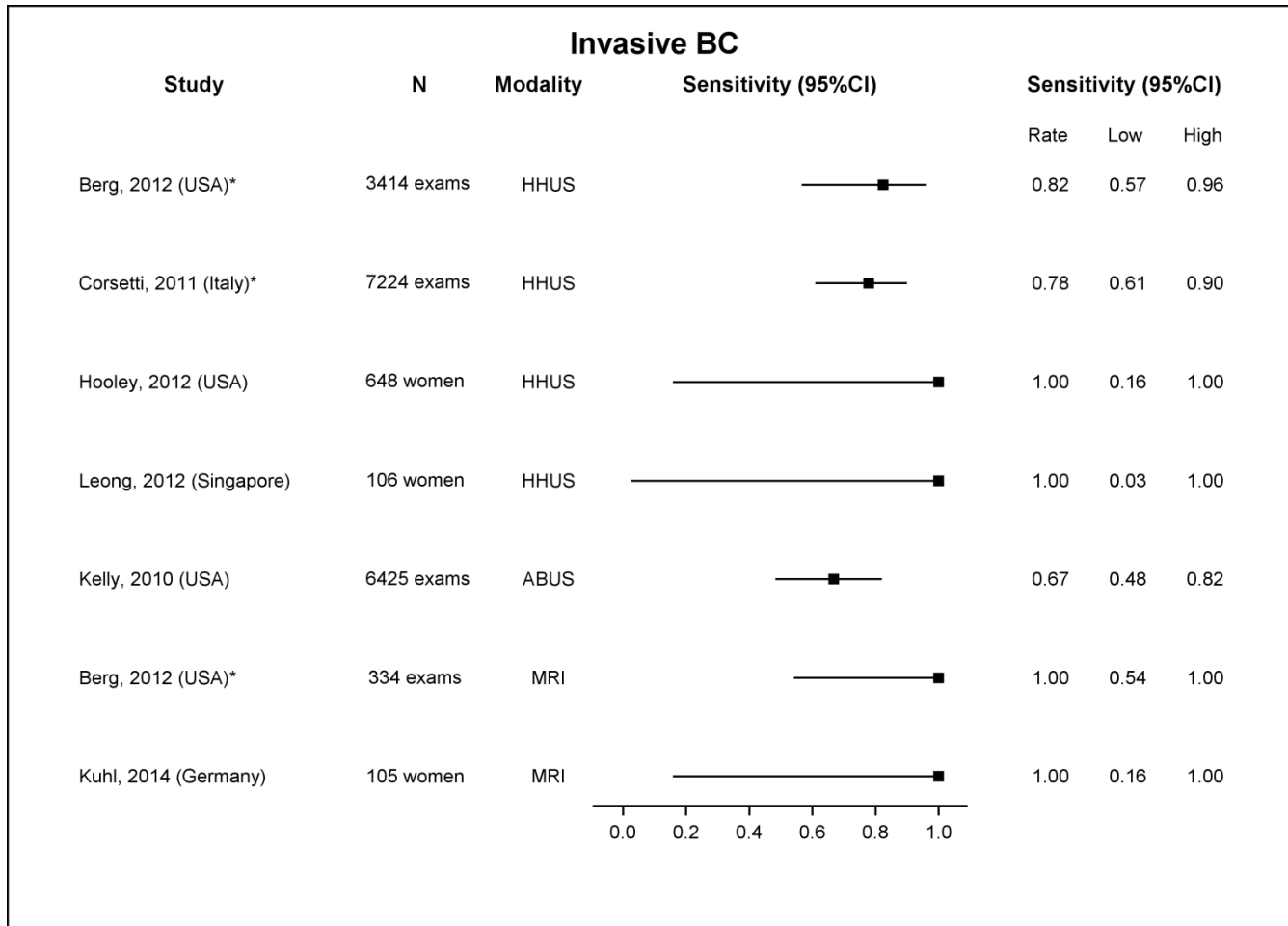


Note: DCIS cases are included as “All Breast Cancer.”

*Good-quality study .

Abbreviations: ABUS=automated whole breast ultrasound; HHUS=hand-held ultrasound.

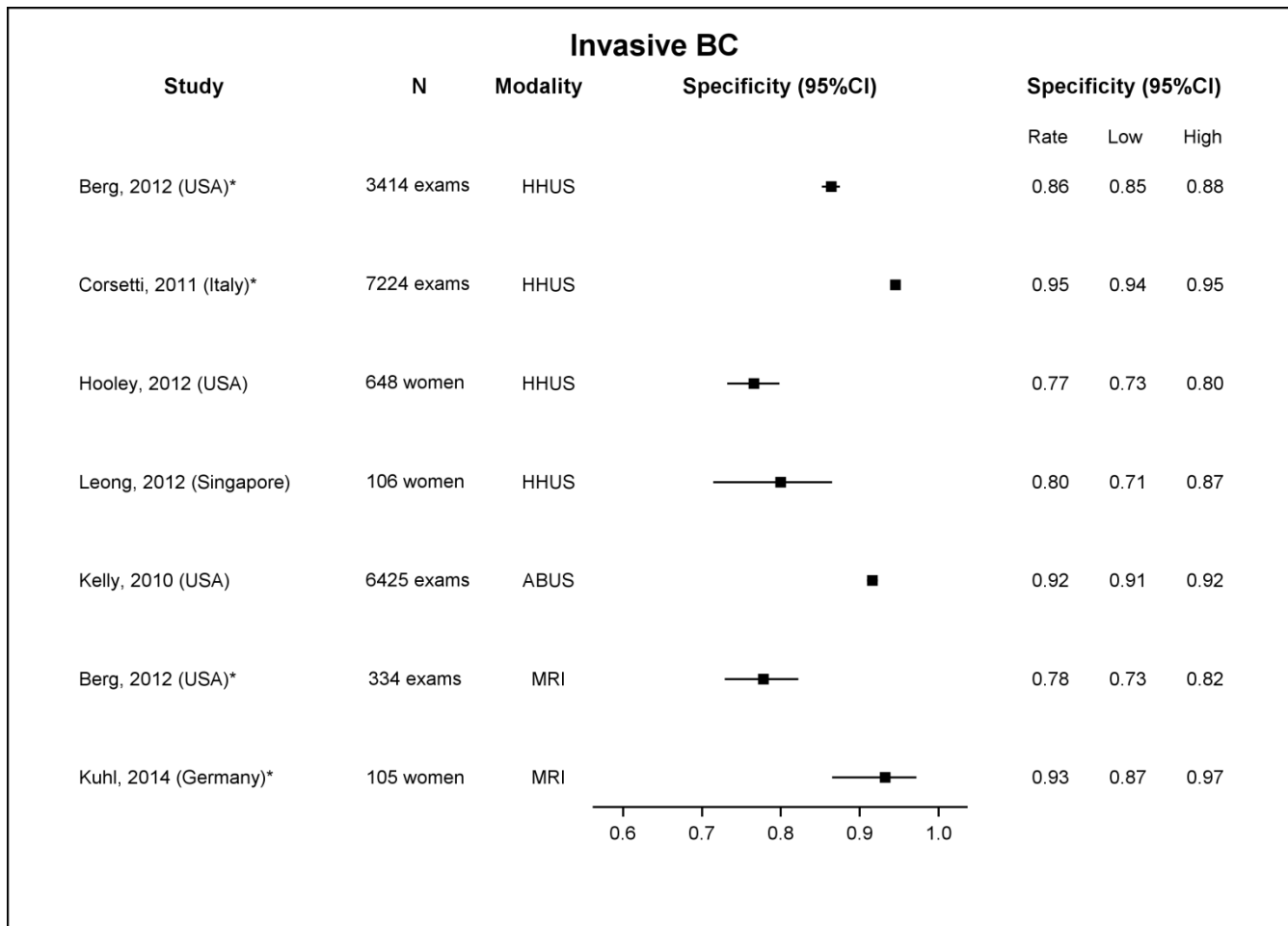
Figure 4. Sensitivity of Supplemental HHUS and MRI in Detecting Invasive Breast Cancer



*Good-quality study.

Abbreviation: HHUS=hand-held ultrasound.

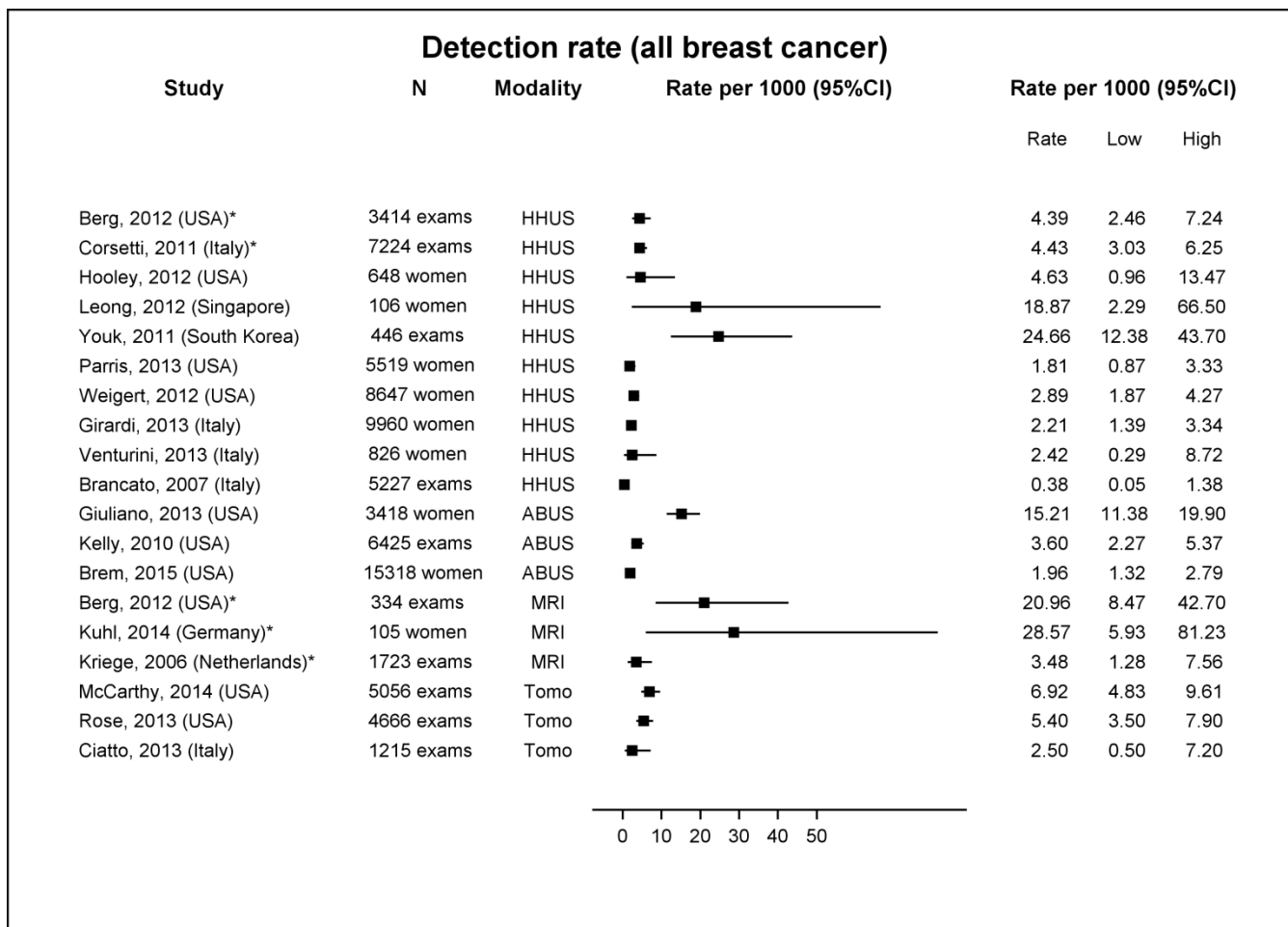
Figure 5. Specificity of Supplemental HHUS and MRI in Detecting Invasive Breast Cancer



*Good-quality study.

Abbreviation: HHUS=hand-held ultrasound.

Figure 6. Breast Cancer Detection Rates of Supplemental HHUS, ABUS, MRI, and DBT

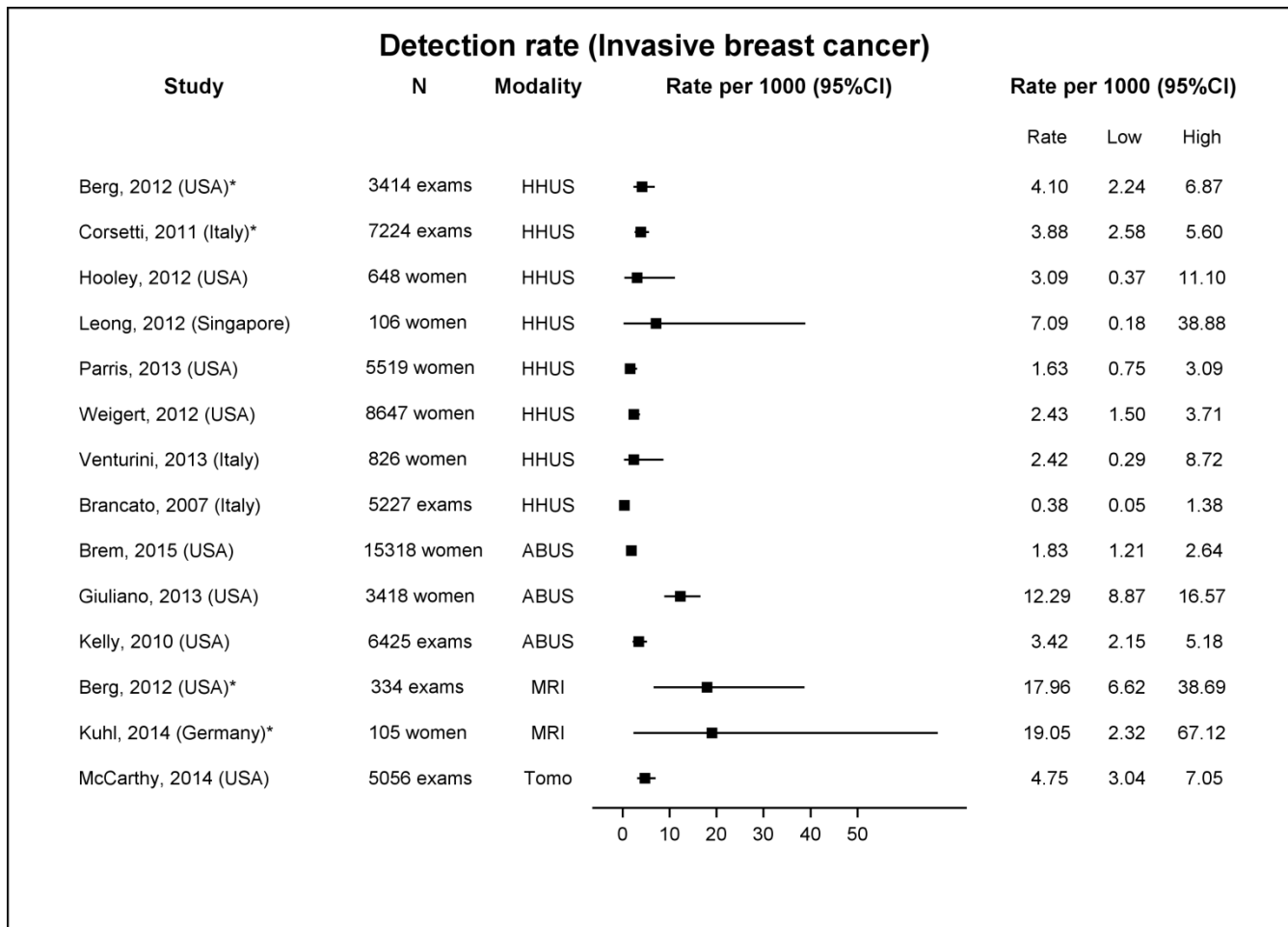


Note: Includes DCIS cases.

*Good-quality study.

Abbreviations: ABUS=automated whole breast ultrasound; DBT=digital breast tomosynthesis; HHUS=hand-held ultrasound.

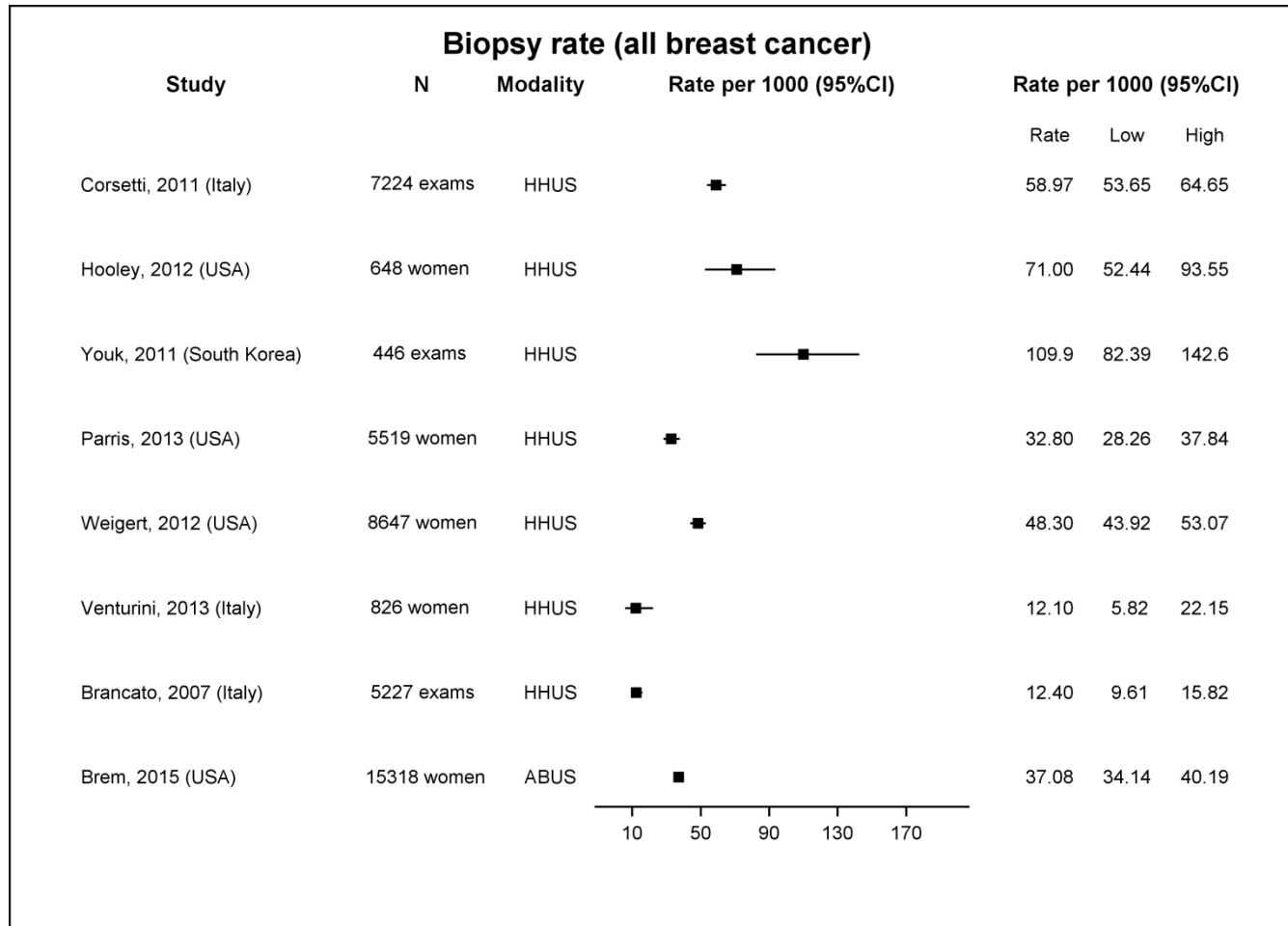
Figure 7. Invasive Breast Cancer Detection Rates of Supplemental HHUS, ABUS, MRI, and DBT



*Good-quality study.

Abbreviations: ABUS=automated whole breast ultrasound; DBT=digital breast tomosynthesis; HHUS=hand-held ultrasound.

Figure 8. Biopsy Rates of Supplemental HHUS and ABUS



Abbreviations: ABUS=automated breast ultrasound; HHUS=hand-held ultrasound.

Table 1. BI-RADS Breast Density Descriptions

2013 BI-RADS Atlas	2003 BI-RADS Atlas	1998 BI-RADS Atlas
a = The breasts are almost entirely fatty	1 = The breast is almost entirely fat (<25% glandular)	1 = The breast is almost entirely fat
b = There are scattered areas of fibroglandular density	2 = There are scattered fibroglandular densities (approximately 25% - 50% glandular)	2 = There are scattered fibroglandular densities
c = The breasts are heterogeneously dense, which may obscure small masses	3 = The breast tissue is heterogeneously dense, which could obscure detection of small masses (approximately 51% - 75% glandular)	3 = The breast tissue is heterogeneously dense. This may lower the sensitivity of mammography
d = The breasts are extremely dense, which lowers the sensitivity of mammography	4 = The breast tissue is extremely dense. This may lower the sensitivity of mammography (>75% glandular)	4 = The breast tissue is extremely, dense, which could obscure a lesion on mammography

Abbreviation: BI-RADS=Breast Imaging-Reporting and Data System.

Table 2. 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 <u>all</u> 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: D.E.N.S.E® State Efforts.

*As of September 2015.

[†] During the 2015 legislative season.

Table 3. U.S. Studies of BI-RADS Density Categories and Consistency of Population Categorization

Study Time between assessments	Exam/ Reading	a* (%)	b† (%)	c‡ (%)	d§ (%)	% of women categorized in the same BI-RADS breast density category at both assessments (4 categories)	Kappa Statistic Interpretation: Intraobserver	Kappa Statistic Interpretation: Interobserver
Harvey, 2013 ⁵⁵ <36 months	Exam 1	9.4	45.2	37.9	7.5	68.1		Interobserver K: 0.49 to 0.56 (Moderate Agreement)
	Exam 2	10.2	45.1	37.2	7.2			
Spayne, 2012 ⁵⁶ 3-24 months	Exam 1	9.8	61.0	26.6	2.5	77.2	Intraobserver K: Weighted, 0.70 (Substantial Agreement)	
	Exam 2	9.2	60.2	28.1	2.5			
Gard, in press ⁵⁷ 6 months	Reading 1	6.1	44.3	38.3	11.4	71.4	Intraobserver K: 0.50 to 0.81 (Moderate to Almost Perfect Agreement)	Interobserver K: 0.02 to 0.72 (Slight to Substantial Agreement)
	Reading 2	4.5	39.2	47.0	9.3			

* 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.

Abbreviation: BI-RADS=Breast Imaging-Reporting and Data System.

Table 4. U.S. Studies of Potential Misclassification of BI-RADS Density Categorization for Women and by Density Categories

Study	% of women receiving a different breast density classification at 2 nd exam (4 categories)	% of women receiving an opposite breast density classification at 2 nd exam (2 categories)	Proportion of dense exams reclassified as non-dense* (c or d) to (a or b)	Proportion of dense exams reclassified as non-dense† (c) to (a or b)	Proportion of dense exams reclassified as non-dense‡ (c) to (b)	Proportion of non-dense exams reclassified as dense§ (a or b) to (c or d)	Proportion of non-dense exams reclassified as dense (b) to (c or d)	Proportion of non-dense exams reclassified as dense¶ (b) to (c)
Harvey, ⁵⁵ 2013	31.9	18.7	0.217	0.208	0.200	0.163	0.160	0.154
Spayne, 2012 ⁵⁶	22.8	12.6	0.191	0.186	0.182	0.100	0.097	0.095
Gard, in press ⁵⁷	28.6	16.9	0.103	0.103	0.103	0.234	0.234	0.228

* Categorized as “heterogeneously dense” or “extremely dense” at first exam and “almost entirely fat” or “scattered fibroglandular densities” at second exam.

† Categorized as “heterogeneously dense” at first exam and “almost entirely fat” or “scattered fibroglandular densities” at second exam.

‡ Categorized as “heterogeneously dense” at first exam and “scattered fibroglandular densities” at second exam.

§ Categorized as “almost entirely fat” or “scattered fibroglandular densities” at first exam and “heterogeneously dense” or “extremely dense” at second exam.

|| Categorized as “scattered fibroglandular densities” at first exam and “heterogeneously dense” or “extremely dense” at second exam.

¶ Categorized as “scattered fibroglandular densities” at first exam and “heterogeneously dense” at second exam.

Abbreviation: BI-RADS=Breast Imaging-Reporting and Data System.

Table 5. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Study and Population Characteristics

Author, Year USPSTF Quality	Location and Study dates	Design and Setting	Study Population	Subgroup Population	Patient Characteristics
Berg, 2012 ⁶¹ Good	United States 1/2004 - 2009	Diagnostic accuracy Multiple academic medical centers	Asymptomatic women with a negative mammogram, dense breasts and elevated risk for breast cancer N=2,262 women (7,473 exams)	Women without a personal history of breast cancer, BRCA 1/2 mutation carriers and women with a history of chest, mediastinal or axillary irradiation N=1,216 women (3,414 exams)	Mean age: 55.2 y* Personal hx of breast cancer: 53.1%* History of chest, mediastinal or axillary irradiation:0.3%* BRCA 1/2 mutation: 0.9%*
Corsetti, 2011 ⁶² Good	Italy 2001 - 2006	Diagnostic accuracy Charity funded breast cancer screening service	Asymptomatic women with a negative mammogram N=5509 women (12,504 exams)	Women with BI-RADS c or d density and a negative mammogram undergoing HHUS N=3,356 women (7,224 exams)	Age distribution: <50 y: 55% ≥50 y: 45%
Hooley, 2012 ⁶³ Fair	Connecticut, United States 10/2009 – 9/2010	Diagnostic accuracy Single academic health system	Women with dense breasts receiving HHUS N=935 women	Women heterogeneous or extremely dense breasts and a negative mammogram N=648 women	Mean age: 52 y* Lifetime risk (based on Gail model) Low/average: 66%* Intermediate: 15.9%* High: 9.35%* Unknown: 9%*
Leong, 2012 ⁶⁴ Fair	Singapore 9/2002 – 11/2004	Diagnostic accuracy General hospital radiology practice	Women with heterogeneous or extremely dense breasts and a negative mammogram N=141 women	Women with heterogeneous or extremely dense breasts and a negative mammogram and complete followup N=106 women	Mean age: 45.1 y Family hx of BC: 20.9% Personal hx of BC: 5%
Youk, 2011 ⁶⁵ Fair	South Korea 7/2001 – 6/2005	Diagnostic accuracy Single academic medical center	Women with negative mammogram and dense breasts N=1,046 women (1,507 exams)	Women with BI-RADS c or d density, no prior history of breast cancer and a negative mammogram N=446 exams	Mean age: 47.5 y* Personal hx of BC: 61.8%*

Table 5. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Study and Population Characteristics

Author, Year USPSTF Quality	Location and Study dates	Design and Setting	Study Population	Subgroup Population	Patient Characteristics
Parris, 2013 ⁷¹ Fair	Connecticut, United States 10/2009 – 9/2010	Cohort Single mammography outpatient facility	Women undergoing HHUS after reporting law passed. Included women with negative mammograms and HHUS findings from women with abnormal mammograms included for breast quadrants in which mammogram was negative N=5,519 women	NA	Mean age: 53.6 y Heterogeneous (>50% density) or very dense breasts: 89% Family hx of BC: 42% Personal hx of BC: 6%
Weigert, 2012 ⁷³ Fair	Connecticut, United States 11/2009 – 11/2010	Cohort 6 radiology practices with 12 sites	Women with 50% or more dense breast tissue and a negative mammogram N=8,647 women (8,647 exams)	NA	Mean age: 54.4 y
Girardi, 2013 ⁷⁰ Fair	Italy	Cohort Single breast clinic	Asymptomatic women with a negative mammogram N=22,131 women	Women with BI-RADS c or d density and a negative N=9,960 women	Mean age: 51.2y* Pre-menopausal: 50%* Personal hx of breast cancer: 9.8%*
Venturini, 2013 ⁷² Fair	Italy	Cohort	Women undergoing screening mammography N=1,666 women	Women with BI-RADS c or d density or implants N=826 women	Age Distribution: 40-44 y: 44.1.1%* 45-49 y: 55.9%* Family hx of BC: 24%* Previous benign biopsy: 3.8% Pre-menopausal: 64%
Brancato, 2007 ⁶⁹ Fair	Italy	Cohort Single large breast screening clinic	Asymptomatic women with BI-RADS c or d density and a negative mammogram N=5,227 women	NA	Age Distribution: <40 y: 4.5% 40-49 y: 68% 50-59 y: 19%

*Data are representative of entire study population, not necessarily the subgroup abstracted.

Abbreviations: BC=breast cancer; BI-RADS=Breast Imaging-Reporting and Data System; HHUS = hand held ultrasound; hx = history; NA=not applicable.

Table 6. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Study and Population Characteristics

Author, Year USPSTF Quality	Location and Study Dates	Design and Setting	Study Population	Subgroup Population	Patient Characteristics
Brem, 2015 ⁶⁴ Fair	United States 2009 – 2011	Cohort Breast screening centers, including academic health systems	Asymptomatic women ages 25 years or older with dense breasts, no previous diagnosis of breast cancer in the past year, no prior breast surgeries or percutaneous procedures in the past year N=15,318 women	N/A	Mean age: 53.3 y Premenopausal: 38.1% Family hx or BC: 44.8% Personal hx of BC: 3.6% BRCA 1/2 carriers: 1% Hormonal therapy (any use): 31.2%
Kelly, 2010 ⁶⁶ Fair	United States 1/2003 – 7/2007	Diagnostic accuracy 8 breast radiology facilities across the United States	Asymptomatic women with either dense breasts, family or personal history, and/or breast implants who had mammography and ABUS N=4,419 women (6,425 exams)	N/A	Mean age: 53 y* Dense Breasts: 68% Family hx of BC: 30%* Personal hx of BC: 10%* Breast Implants: 11%* BRCA 1/2 carriers: 0.1%*
Giuliano, 2013 ⁶⁷ Fair	United States 1/2009 – 5/2011	Cohort Academic medical center	Women presenting for breast cancer screening N=7,497 women	Women with >50% breast density (Wolf classification) who received ABUS in addition to mammography N=3,418 women	"...no pre-existing predictors of breast cancer, such as personal or family history or BRCA gene positive."

*Data are representative of entire study population, not necessarily the subgroup abstracted.

Abbreviations: ABUS=automated whole breast ultrasound; BC=breast cancer; BI-RADS=Breast Imaging-Reporting and Data System; CG=control group; IG=intervention group.

Table 7. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Study and Population Characteristics

Author, Year USPSTF Quality	Location and Study Dates	Design and Setting	Study Population	Subgroup Population	Patient Characteristics
Berg, 2012 ⁶¹ Good	United States 4/2004 - 2009	Diagnostic accuracy Multiple academic medical centers	Asymptomatic women with a negative mammogram, dense breasts and elevated risk for breast cancer who underwent 3 rounds of ultrasound screening and 1 round of MRI screening N=612 exams	Women without a personal history of breast cancer, BRCA 1/2 mutation carriers and women with a history of chest, mediastinal or axillary irradiation. N=334 exams	Mean age: 56.8 y Dense breasts: 100% BRCA 1/2 mutation: 0.5-0.8%*
Kuhl, 2014 ²³ Good	Germany 1/2009 – 6/2010	Diagnostic accuracy One university hospital	Women with a negative mammogram at high risk for breast cancer (based on family and/or personal history) (443 women)	Women with BI-RADS c or d density without family or personal history of breast cancer (105 women)	Mean age: 53.2 y
Kriege, 2006 ⁶⁸ Good	Netherlands 11/1999 – 10/2003	Diagnostic accuracy Six familial cancer clinics	Women aged 25-70 years with a lifetime cancer risk > 15% (1,779 women/4,134 exams)	Women with BI-RADS c or d dense breasts who were not BCRA mutation carriers (1,723 exams)	Mean age: 40 y* Family hx of BC: 100%* Pre-menopausal: 76.6%* Using hormonal therapy: 4%*

*Data are representative of entire study population, not necessarily the subgroup abstracted.

Abbreviations: BC=breast cancer; BI-RADS=Breast Imaging-Reporting and Data System; US=ultrasound.

Table 8. Supplemental DBT Screening for Breast Cancer in Women With Dense Breasts: Study and Population Characteristics

Author, Year USPSTF Quality	Location and Study Dates	Design and Setting	Study Population	Subgroup Population	Patient Characteristics
McCarthy, 2014 ⁷⁵ Fair	Philadelphia, PA 9/2010 – 2/2013	Cohort University health system	Asymptomatic women, including those with implants and large breasts requiring multiple images N=18,220 women	Asymptomatic women with dense breasts DBT+DM (IG) = 5,056 exams DM only (CG) = 3,489 exams	Age: 70% of patients ages 50 years and older
Haas, 2013 ⁷⁶ Fair	New Haven, Connecticut 10/2011 – 9/2012	Cohort Outpatient radiology clinics, mobile mammography van, tertiary care hospital	Women presenting for screening, including women with a personal history of breast cancer, undergoing DBT plus mammography DBT+DM (IG) = 6,100 women DM only (CG) = 7,058 women	Asymptomatic women with dense breasts N=4,794 exams	Dense Breasts (c or d): 43.2%* Age:* <40 y: 2.7% 40-49 y: 30.6% 50-59 y: 33.3% 60-69 y: 23.3% >70 y: 10.1% Personal history of BC: 5.5%*
Rose, 2013 ⁷⁷ Fair	Houston, TX 2010 – 2012	Retrospective cohort One cancer center	Asymptomatic women DBT+DM (IG) = 9,499 exams DM only (CG) = 13,856 exams	Asymptomatic women with BI-RADS 3-4 density DBT+DM (IG) = 4,666 exams DM only (CG) = 7,009 exams	IG:* Mean age: 54.5 y <50 y: 37.5% 50-64 y: 46.8% >64 y: 17.5% CG: Mean age: 53.8 y <50 y: 38.5% 50-64 y: 45.4% >64 y: 16.1%
Ciatto, 2013 ⁷⁰ Fair	Italy 9/2011 – 6/2012	Prospective cohort Multiple breast screening programs	Asymptomatic women undergoing digital mammography followed by DBT N=7,294 women	Asymptomatic women with BI-RADS 3-4 density N=1,215 women	Median age: 58 y* Age range: 48 – 71 y*

*Data are representative of entire study population, not necessarily the subgroup abstracted.

Abbreviations: BC=breast cancer; BI-RADS=Breast Imaging-Reporting and Data System; CG=control group; DBT=digital breast tomosynthesis; DM=digital mammography; IG=intervention group.

Table 9. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Screening Test Characteristics

Author, Year USPSTF Quality	Supplemental Test and Comparison	# Screening Rounds (per woman)	Followup Period and LTFU	Readers	Reference Standard
Berg, 2012 ⁶¹ Good	Negative mammogram followed by HHUS	3	12+ months LTFU < 3%	Ultrasounds done by technologists and read by radiologists across 12 sites	Biopsy and/or one year clinical followup
Corsetti, 2011 ⁶² Good	Negative mammogram followed by HHUS in women with BI-RADS 3-4 breast density	2 (mean)	12 months LTFU estimated at <5%	6 radiologists; single read	Biopsy or 1-year followup with search of all hospital registries in catchment region for breast cancer cases
Hooley, 2012 ⁶³ Fair	Negative mammogram followed by HHUS	1-2 per woman	15+ months 19% LTFU	8 breast radiologists with 2-32 years' experience	Biopsy or a followup mammogram with or without US within 15 months
Leong, 2012 ⁶⁴ Fair	Negative mammogram followed by HHUS	NR	12-24 months 35 women (25%) LTFU	3 sonographers with 4-12 years of US experience performed the exams. 4 radiologists with 3-15 years of breast US experience reviewed the exams	Biopsy or chart review if negative assessment at 2 years
Youk, 2011 ⁶⁵ Fair	Negative mammogram followed by HHUS	1-4	24 months (Excluded patients without followup)	9 radiologists with fellowships or 408 years clinical experience in breast imaging	Biopsy or documented 2-year followup
Parris, 2013 ⁷¹ Fair	Negative mammogram followed by HHUS	1	NR	6 board certified radiologists with 4-32 years' experience	No comprehensively applied reference standard
Weigert, 2012 ⁷³ Fair	Negative mammogram followed by HHUS	1 per woman	NR	Ultrasounds done by technologists and read by radiologists across 12 sites	No comprehensively applied reference standard
Girardi, 2013 ⁷⁰ Fair	Negative mammogram followed by HHUS	1 per woman	12 months LTFU NR	5 radiologists 6-34 years' experience in breast ultrasound	No comprehensively applied reference standard
Venturini, 2013 ⁷² Fair	Negative mammogram followed by HHUS	1 per woman	NR	4 readers with 10-20 years' experience	No comprehensively applied reference standard
Brancato, 2007 ⁶⁹ Fair	Negative mammogram followed by HHUS (within one month)	NR	NR	Radiologists with 3-15 years breast US experience and at least 500 US performed annually	No comprehensively applied reference standard

Abbreviations: HHUS=hand-held ultrasound; LTFU=lost to followup; NR=not reported; US=ultrasound.

Table 10. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Screening Test Characteristics

Author, Year USPSTF Quality	Supplemental Test and Comparison	# Screening Rounds (per woman)	Followup Period and LTFU	Readers	Reference Standard
Brem, 2015 ⁶⁴ Fair	Mammography with ABUS compared to mammography alone	1 per woman	12 months 12% LTFU	39 radiologists at 11 sites	No comprehensively applied reference standard
Kelly, 2010 ⁶⁶ Fair	Mammography with ABUS compared to mammography alone	1 or more per woman	12 months 20% LTFU	10 radiologists with at least ten years' experience in breast US	Biopsy or one year followup with mammography
Giuliano, 2013 ⁶⁷ Fair	Digital mammography with ABUS (IG) compared to digital mammography alone (CG)	1 per woman	12 months LTFU not reported	2 radiologists with more than ten years' experience in breast ultrasound and two years' experience in ABUS. All images double read	No comprehensively applied reference standard

Abbreviations: ABUS=automated whole breast ultrasound; LTFU=lost to followup; NR=not reported.

Table 11. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Screening Test Characteristics

Author, Year USPSTF Quality	Supplemental Test and Comparison	# Screening Rounds (per woman)	Followup Period and LTFU	Readers	Reference Standard
Berg, 2012 ⁶¹ Good	3 rounds of negative mammogram and ultrasound followed by MRI	1	12 months 15 women LTFU	Multiple radiologists who completed skills testing for mammographic and ultrasound interpretation	Biopsy and/or one year clinical followup
Kuhl, 2014 ²³ Good	Negative mammogram followed by MRI	1-2	24 months 2% LTFU overall, 0% in subgroup	2 readers with 6-18 years' experience with an annual caseload of approximately 800 MRI images	Biopsy of positive results and two years of followup for negative results
Kriege, 2006 ⁶⁸ Good	Negative mammogram followed by MRI	2.3 (mean)*	12 months	NR	Biopsy and one year followup for negative results

*Data are representative of entire study population, not necessarily the subgroup abstracted.

Abbreviations: LTFU=lost to followup; NR=not reported.

Table 12. Supplemental DBT Screening for Breast Cancer in Women With Dense Breasts: Screening Test Characteristics

Author, Year USPSTF Quality	Supplemental Test and Comparison	# Screening Rounds (per woman)	Followup Period and LTFU	Readers	Reference Standard
McCarthy, 2014 ⁷⁵ Fair	Digital mammography alone compared to digital mammography and DBT	1-2	No followup except on biopsy results	6 radiologists with 3-22 years' experience and formal training in DBT	No comprehensively applied reference standard
Haas, 2013 ⁷⁶ Fair	Digital mammography alone compared to digital mammography and DBT	NR	No followup except on biopsy results	8 radiologists with 2-23 years' experience and certification in DBT	No comprehensively applied reference standard
Rose, 2013 ⁷⁷ Fair	Digital mammography alone (CG) compared to digital mammography and DBT (IG)	1	No followup except on biopsy results	6 radiologists with 2-32 years' experience	No comprehensively applied reference standard
Ciatto, 2013 ⁷⁰ Fair	Digital mammography followed by DBT	1	No followup except on biopsy results	8 radiologists with 3-13 years' experience in mammography and basic training in DBT	No comprehensively applied reference standard

Abbreviations: CG=control group; DBT=digital breast tomosynthesis; IG=intervention group; LTFU=lost to followup.

Table 13. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Test Performance Characteristics

Study, Year USPSTF Quality	Study N	Subgroup N	Study/Subgroup characteristics	Breast Cancer Type	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value	Negative Predictive Value
Berg, 2012 ⁶¹ Good	2,262 women 7,473 exams	1,216 women	Mean age: 55.2*	All Breast Cancer	0.833 (0.586 to 0.964)	0.864 (0.852 to 0.875)	0.032	0.998
		3,414 exams	Personal hx of BC: 53.1%* Hx of chest, mediastinal or axillary irradiation: 0.3% BRCA 1/2 mutation: 0.9%*	Invasive Breast Cancer	0.824 (0.566 to 0.962)	0.864 (0.852 to 0.875)	0.030	0.999
Corsetti, 2011 ⁶² Good	5,509 women 12,504 exams	3,356 women	Age distribution: <50 years: 55% ≥50 years: 45%	All Breast Cancer	0.800 (0.645 to 0.910)	0.945 (0.940 to 0.950)	0.075	0.999
		7,224 exams		Invasive Breast Cancer	0.778 (0.609 to 0.899)	0.945 (0.940 to 0.950)	0.066	0.999
Hooley, 2012 ⁶³ Fair	935 women	648 women	Mean age: 52 y* Lifetime risk (based on Gail model) Low/average: 66%* Intermediate: 15.9%* High: 9.35%* Unknown: 9%*	All Breast Cancer	1.000 (0.292 to 1.000)	0.767 (0.733 to 0.798)	0.020	1.000
				Invasive Breast Cancer	1.000 (0.158 to 1.000)	0.766 (0.732 to 0.798)	0.013	1.000
Leong, 2012 ⁶⁴ Fair	141 women	106 women	Mean age: 45.1 y Family hx of BC: 20.9% Personal hx of BC: 5%	All Breast Cancer	1.000 (0.158 to 1.000)	0.789 (0.697 to 0.862)	0.083	1.000
				Invasive Breast Cancer	1.000 (0.025 to 1.000)	0.800 (0.714 to 0.865)	0.045	1.000
Youk, 2011 ⁶⁵ Fair	1,046 women 1,507 exams	446 exams	Mean age: 47.5 y* Personal history of BC: 61.8%*	All Breast Cancer	1.000 (0.715 to 1.000)	0.717 (0.672 to 0.759)	0.082	1.000

*Data are representative of entire study population, not the subgroup with dense breasts abstracted.

Abbreviations: BC=breast cancer; BI-RADS=Breast Imaging-Reporting and Data System; HHUS=hand-held ultrasound; hx=history.

Table 14. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Test Performance Characteristics

Author, Year Quality	Study N	Subgroup N	Study/Subgroup characteristics	Breast Cancer Type	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value	Negative Predictive Value
Kelly, 2010 ^{6b} Fair	4,419 women 6,425 exams	NA	Mean age: 53 y* Dense breasts: 68% Family hx of BC: 30%* Personal hx of BC: 10%* Breast implants: 11%* BRCA 1/2 carriers: 0.1%*	All Breast Cancer	0.676 (0.495 to 0.826)	0.916 (0.910 to 0.923)	0.041	0.998
				Invasive Breast Cancer	0.667 (0.482 to 0.820)	0.916 (0.909-0.923)	0.039	0.998

*Data are representative of entire study population, not the subgroup with dense breasts abstracted.

Abbreviations: ABUS=automated whole breast ultrasound; BC=breast cancer; hx=history; NA=not applicable.

Table 15. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Test Performance Characteristics

Study, Year USPSTF Quality	Study N	Subgroup N	Study/Subgroup characteristics	Breast Cancer Type	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value	Negative Predictive Value
Berg, 2012 ⁶¹ Good	2,262 women	334 exams	Mean age: 56.8 y Dense breasts: 100% BRCA 1/2 mutation: 0.5%-0.8%*	All Breast Cancer	1.000 (0.590 to 1.000)	0.781 (0.732 to 0.825)	0.090	1.000
	7,473 exams			Invasive Breast Cancer	1.000 (0.541 to 1.000)	0.788 (0.729 to 0.822)	0.077	1.000
Kuhl, 2014 ²³ Good	443 women	105 women	Mean age: 53.2 y	All Breast Cancer	1.000 (0.292 to 1.000)	0.941 (0.876 to 0.978)	0.333	1.000
				Invasive Breast Cancer	1.000 (0.158 to 1.000)	0.932 (0.865 to 0.972)	0.222	1.000
Kriege, 2006 ⁶⁸ Good	1,779 women 4,134 exams	1,723 exams	Mean age: 40 y* Dense breasts; 100% Family hx of BC: 100%*	All Breast Cancer	0.750 (0.349 to 0.968)	0.887 (0.872 to 0.902)	0.030	0.999

*Data are representative of entire study population, not the subgroup with dense breasts abstracted.

Abbreviations: BC=breast cancer; hx=history.

Table 16. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Author, Year USPSTF Quality	Study N	Subgroup N	Breast Cancer Type	Cancers Detected	Cancer Detection Rate (95% CI)	Recall/Biopsy Rate* (95% CI)
Berg, 2012 ⁶¹ Good	2,262 women 7,473 exams	1,216 women 3,414 exams	All Breast Cancer	15/3,414 exams	4.4 per 1,000 exams (2.5 to 7.2)	Recall: 139 per 1,000 exams (127.7 to 151.2)
			Invasive Breast Cancer	14/3,414 exams	4.1 per 1,000 exams (2.2 to 6.9)	
			Interval Cancers	3/3,414 exams	0.9 per 1,000 exams (0.2 to 2.6)	
Corsetti, 2011 ⁶² Good	5,509 women 12,504 exams	3,356 women 7,224 exams	All Breast Cancer	32/7,224 exams	4.4 per 1,000 exams (3.0 to 6.2)	Biopsy: 59 per 1,000 exams (53.7 to 64.7)
			Invasive Breast Cancer	28/7,224 exams	3.9 per 1,000 exams (2.6 to 5.6)	
			Interval Breast Cancer	8/7,224 exams	1.1 per 1,000 exams (0.5 to 2.2)	
Hooley, 2012 ⁶³ Fair	935 women	648 women	All Breast Cancer	3/648 women	4.6 per 1,000 women (1.0 to 13.5)	Recall: 236 per 1,000 women (203.9 to 270.4) Biopsy: 71 per 1,000 women (52.4 to 93.6)
			Invasive Breast Cancer	2/648 women	3.1 per 1,000 women (0.4 to 11.1)	
Leong, 2012 ⁶⁴ Fair	141 women	106 women	All Breast Cancer	2/106 women	18.9 per 1,000 women (1.7 to 50.3)	Recall: 170 per 1,000 women [†] Biopsy: 132 per 1,000 women (74.1 to 211.7) [†]
			Invasive Breast Cancer	1/106 women	9.4 per 1,000 women (0.2 to 51.4)	
			Interval Breast Cancer	0/106	0 per 1,000 women	
Youk, 2011 ⁶⁵ Fair	1,046 women 1,507 exams	446 exams	All Breast Cancer	11/446 exams	24.7 per 1,000 exams (12.4 to 43.7)	Recall: 137 per 1,000 exams (106.3 to 172.2) Biopsy: 110 per 1,000 exams (82.4 to 142.6)
Parris, 2013 ⁷¹ Fair	5,519 women	NA	All Breast Cancer	10/5,519 women	1.8 per 1,000 women (0.9 to 3.3)	Recall: NR Biopsy: 33 per 1,000 women (28.3 to 37.8)
			Invasive Breast Cancer	9/5,519 women	1.6 per 1,000 women (0.7 to 3.1)	
Weigert, 2012 ⁷³ Fair	8,647 women 8,647 exams	NA	All Breast Cancer	25/8,647 women	2.9 per 1,000 women (1.9 to 4.3)	Recall: 138 per 1,000 women (130.8 to 145.4) Biopsy: 48 per 1,000 women (43.9 to 53.1)
			Invasive Breast Cancer	21/8,647 women	2.4 per 1,000 women (1.5 to 3.7)	
Girardi, 2013 ⁷⁰ Fair	22,131 women	9,960 women	All Breast Cancer	22/9,960 women	2.2 per 1,000 women (1.4 to 3.3)	Recall : NR Biopsy : NR

Table 16. Supplemental HHUS Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Author, Year USPSTF Quality	Study N	Subgroup N	Breast Cancer Type	Cancers Detected	Cancer Detection Rate (95% CI)	Recall/Biopsy Rate* (95% CI)
Venturini, 2013 ⁷² Fair	1,666 women	826 women	All Breast Cancer	2/826 women	2.4 per 1,000 women (0.3 to 8.7)	Recall: 95 per 1,000 women (75.4 to 116.5) Biopsy: 12 per 1,000 women (5.8 to 22.2)
			Invasive Breast Cancer	2/826 women	2.4 per 1,000 women (0.3 to 8.7)	
Brancato, 2007 ⁶⁹ Fair	5,227 women	NA	All Breast Cancer	2/5,227 women	0.4 per 1,000 women (0 to 1.4)	Recall: 21 per 1,000 women (17.3 to 25.3) Biopsy: 12 per 1,000 women (9.6 to 15.8)
			Invasive Breast Cancer	2/5,227 women	0.4 per 1,000 women (0 to 1.4)	

*Biopsy rate includes needle aspiration, core needle, and open biopsies.

†Data are based on the 106 women with complete followup (out of 141 total).

Abbreviations: CI=confidence interval; HHUS=hand-held ultrasound; NA=not applicable; NR=not reported.

Table 17. Supplemental ABUS Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Author, Year USPSTF Quality	Study N	Subgroup N	Breast Cancer Type	Cancers Detected	Cancer Detection Rate (95% CI)	Recall/ Biopsy Rate* (95% CI)
Brem, 2015 ⁷⁴ Fair	15,318 women	NA	All Breast Cancer	30/15318 women	1.9 per 1,000 exams (1.3 to 2.8)	Recall: 135 per 1,000 exams (129.3 to 140.2) Biopsy: 36 per 1,000 exams (34.1 to 40.2)
			Invasive Breast Cancer	28/15318 women	1.8 per 1,000 exams (1.2 to 2.6)	
Kelly, 2010 ⁶⁶ Fair	4,419 women 6,425 exams	NA	All Breast Cancer	23/6425 exams	3.6 per 1,000 exams (2.3 to 5.4)	Recall: 87 per 1,000 exams (80.2 to 94.2)
			Invasive Breast Cancer	22/6425 exams	3.4 per 1,000 exams (2.1 to 5.2)	
			Interval Breast Cancers	11/6425 exams	1.7 per 1,000 exams (0.9 to 3.1)	
Giuliano, 2013 ⁶⁷ Fair	7,497 women DM only (CG): 4,076 women	DM+ABUS (IG): 3,418 women	All Breast Cancer	IG: 52/3418 women CG: 19/4076 women	IG: 15.21 per 1,000 women (11.4 to 19.9) CG: 4.7 per 1,000 women (2.8 to 7.3)	Recall: 15 per 1,000 women (11.1 to 19.6)
			Invasive Breast Cancer	IG: 42/3418 women CG: 19/4076 women	IG: 12.3 per 1000 women (8.9 to 16.6) CG: 4.7 per 1,000 women (2.8 to 7.3)	

*Biopsy rate includes needle aspiration, core needle, and open biopsies.

Abbreviations: ABUS=automated whole breast ultrasound; BC=breast cancer; BI-RADS=Breast Imaging-Reporting Data System; CG=control group; CI=confidence interval; DM=digital mammography; IG=intervention group.

Table 18. Supplemental MRI Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Author, Year USPSTF Quality	Study N	Subgroup N	Breast Cancer Type	Cancers Detected	Cancer Detection Rate (95% CI)	Recall/Biopsy Rate* (95% CI)
Berg, 2012 ⁶¹ Good	2,262 women 7,473 exams	334 exams	All Breast Cancer	7/334 exams	21 per 1,000 exams (8.5 to 42.7)	Recall: 234 per 1,000 exams (189.2 to 282.7)
			Invasive Breast Cancer	6/334 exams	18 per 1,000 exams (6.6 to 38.7)	
Kuhl, 2014 ²³ Good	443 women	105 women	All Breast Cancer	3/105 women	28.6 per 1,000 women (5.9 to 81.2)	Recall: 86 per 1,000 women (40.0 to 156.5)
			Invasive Breast Cancer	2/105 women	19 per 1,000 women (2.3 to 67.1)	
Kriege, 2006 ⁶⁸ Good	1,779 women 4,134 exams	1,723 exams	All Breast Cancer	6/1723 exams	3.5 per 1,000 exams (1.3 to 7.6)	Recall: 115 per 1,000 exams (100.2 to 130.9)
			Interval Breast Cancer	2/1723 exams	1.2 per, 1000 exams (0.1 to 4.2)	

*Biopsy rate includes needle aspiration, core needle, and open biopsies.

Abbreviations: CI=confidence interval; DCIS=ductal carcinoma in situ; NR=not reported.

Table 19. Supplemental DBT Screening for Breast Cancer in Women With Dense Breasts: Cancer Detection Outcomes

Author, Year USPSTF Quality	Study N	Subgroup N	Breast Cancer Type	Cancers Detected	Cancer Detection Rate (95% CI)	Recall/Biopsy Rate* (95% CI)
McCarthy, 2014 ⁷⁵ Fair	18,220 women	DBT+DM (IG) = 5,056 exams DM only (CG) = 3,489 exams	All Breast Cancer	IG: 35/5056 exams CG: 18/3489 exams	IG: 6.9 per 1,000 exams (4.8 to 9.6) CG: 5.2 per 1,000 exams (3.1 to 8.1)	Recall: IG: 108 per 1,000 exams (99.6 to 116.9) CG: 128 per 1,000 exams (117.2 to 139.7)
			Invasive Breast Cancer	IG: 24/5056 exams CG: 12/3489 exams	IG: 4.7 per 1,000 exams (3.0 to 7.0) CG: 3.4 per 1,000 exams (1.8 to 6.0)	
Haas, 2013 ⁶ Fair	DBT+DM (IG) = 6,100 women DM only (CG) = 7,058 women	DBT+DM (IG) = 2,639 exams DM only (CG) = 2,158 exams	All Breast Cancer	NR	NR	Recall: IG: 97 per 1,000 women (86.0 to 108.9) CG: 166 per 1,000 women (150.4 to 182.3)
Rose, 2013 ⁷⁷ Fair	DBT+DM (IG) = 9,499 exams DM only (CG) = 13,856 exams	DBT+DM (IG) = 4,666 exams DM only (CG) = 7,009 exams	All Breast Cancer	IG: 25/4666 exams CG: 28/7009 exams	IG: 5.4 per 1,000 exams (3.5 to 7.9) CG: 4.0 per 1,000 exams (2.7 to 5.8)	Recall: IG: 69 per 1,000 exams (61.9 to 76.7) CG: 91 per 1,000 exams (84.4 to 110.1)
Ciatto, 2013 ⁷⁰ Fair	7,294 exams	1,215 exams	All Breast Cancer	DBT+DM: 8/1215 exams DM only: 5/1215 exams	DBT+DM: 6.6 per 1,000 exams (2.9 to 12.9) DM only: 4.1 per 1,000 exams (1.3 to 9.6)	Recall: DBT+DM: 66 per 1,000 exams (52.6 to 81.3)) DM only: 72 per 1,000 exams (57.6 to 87.6)

*Biopsy rate includes needle aspiration, core needle, and open biopsies.

Abbreviations: CI=confidence interval; CG=control group; DM=digital mammography; IG=intervention group.

Table 20. Cancer Detection, Recall, and Biopsy Rates of Good-Quality Supplemental Screening Studies

Author, Year USPSTF Quality	Country	Modality	Breast Cancer Detection Rate (95% CI)	Invasive Breast Cancer Detection	Recall Rate	Biopsy Rate*
Berg, 2012 ⁶¹ Good	United States	HHUS	4.4 per 1,000 exams (2.5 to 7.2)	93.3% (14/15)	138.8 per 1,000 exams (474/3414)	NR
Corsetti, 2011 ⁶² Good	Italy	HHUS	4.4 per 1,000 exams (3.0 to 6.2)	87.5% (28/32)	NR	59.0 per 1,000 exams (427/7224)
Berg, 2012 ⁶¹ Good	United States	MRI	21.0 per 1,000 exams (8.5 to 42.7)	85.7% (6/7)	233.5 per 1,000 exams (78/334)	NR
Kuhl, 2014 ²³ Good	Germany	MRI	28.6 per 1,000 women (5.9 to 81.2)	66.7% (2/3)	85.7 per 1,000 women (9/105)	NR
Kriege, 2006 ⁶⁸ Good	Netherlands	MRI	3.5 per 1,000 exams (2.0 to 9.1)	NR	109 per 1,000 exams (199/1458)	NR

*Includes needle aspiration, core needle, and open biopsies.

Abbreviations: HHUS=hand-held ultrasound; NR=not reported.

Table 21. Harms Associated With Breast Density Notification: Study and Population Characteristics

Author, Year USPSTF Quality	Country	Design	Study Period	Intervention	N Analyzed	Mean age, y	% Dense Breasts	Population Characteristics
Bottorff, 2007 ⁸⁰ Good	Canada	RCT	2/2002 – 8/2003	IG: letter notification of mammograms results with info on breast density and pamphlet on breast cancer risk factors CG: Usual letter notification of mammography results	Total: 618 IG: 285 CG: 333	IG: 66.1 CG: 65.9	IG: 100% CG: 100%	Screening population 1 st degree relative with BC: IG: 15.7% CG: 18.3% Gail Lifetime Risk: IG: 8.7% CG: 9.1%

Abbreviations: BC=breast cancer; CG=control group; IG=intervention group; RCT=randomized, controlled trial.

Table 22. Harms Associated With Breast Density Notification: Knowledge, Perceived Risk, and Breast Cancer Worry

Study, Year USPSTF Quality	Mean Followup (weeks)	N Analyzed	Treatment Group	Knowledge of breast density as a risk factor	Higher Perceived Lifetime Risk*	Lower Perceived Lifetime Risk Relative to Other Women [†]	Psychological Distress: Worry Often About Breast Cancer [‡]
Bottorff, 2007 ⁸⁰	4	265	IG	85.3%	5.5%	10.5% [§]	23.0%
		314	CG	66.4%	3.7%	15.5% [§]	20.4%
Good	24	270	IG	89.2%	4.1%	10.6%	19.3%
		314	CG	63.8%	2.6%	16.7%	20.9%

*Women who rated their own lifetime risk of breast cancer as “high.”

[†] Women who rated their perceived lifetime risk as “a lot lower” relative to other women.

[‡] Women who responded with “often” when asked how often they worry about getting breast cancer.

[§] Statistically significant difference at $p < 0.05$.

Abbreviations: CG=control group; IG=intervention group.

Table 23. Summary of the Evidence

Key Question(s)	# of Studies (k)	Design	Major Limitations	Consistency	Applicability	Overall Evidence	Summary of Findings
KQ 1. What is the accuracy and inter-rater reliability of BI-RADS determination of breast density?	k=5	Registry-based studies; test sets	Community practice data may be criticized for representing older practice Some of the reclassifications found in the registry data may appropriately reflect actual physiological changes reflected in breast density, however it is unlikely given selection criteria and relatively short time frames for re-examination	Relatively consistent	Two studies were conducted in Spain and Italy One U.S.-based study was conducted within a single healthcare system, so results may underestimate variability in a more diverse population of radiologists	2 good; 3 fair Overall: Fair to good	Results suggest that 1 in 5 women would be categorized into a different BI-RADS density category (dense to non-dense) by the same radiologist after the second screening exam, while 1 in 3 would be recategorized after a second exam read by a different radiologist Reducing variability through the use of double reading with consensus, along with introducing standards and quality measures, may minimize possible harms associated with variation in breast density categorizations
KQ 2. What are the test performance characteristics of newer technologies for breast cancer screening when used as supplemental tests after a negative screening mammography exam in women found to have dense breasts and how do these performance characteristics differ by age and risk factors?	HHUS k=5	Test performance	Most studies not exclusively screening populations Only 1 good-quality study from the US Some had unclear followup protocols for negative results	2 good-quality studies gave more consistent estimates of sensitivity and specificity, fair-quality studies had more variability	HHUS is somewhat operator dependent which may limit generalizability of results	2 good; 3 fair Overall: fair	From good-quality studies for invasive breast cancer: Sens.: 77.8 – 82.4% Spec.: 86.4 – 94.5% PPV: 3 – 7.5% NPV: 99.8 – 99.9%
	ABUS k=1	Test performance	Only 1 study, 20% loss to followup Study authors had financial conflict of interest	N/A	Multi-site United States study	1 fair Overall: Insufficient	All breast cancer: Sens: 67.6% Spec: 91.6% PPV: 4.1% NPV: 99.8%
	MRI k=3	Test performance	Only 1 good-quality study from the US. Study populations were high risk beyond breast density, even with subgroup analysis family history still an important risk factor	Good-quality studies Small numbers, variable findings	2 European studies, 1 from United States, all focused on high risk women	3 good but small studies of high risk women Overall: fair	From 2 good-quality studies for invasive breast cancer: Sens: 100% Spec: 78.1 – 93.2% PPV: 7.7 – 22.2% NPV: 100%

Table 23. Summary of the Evidence

Key Question(s)	# of Studies (k)	Design	Major Limitations	Consistency	Applicability	Overall Evidence	Summary of Findings
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 ultrasound, MRI, or DBT on proximate clinical outcomes, including cancer detection rates, DCIS detection rates, stage at diagnosis, recall rates, biopsy rates, and interval cancer rates?	HHUS k=10	Test performance; cohort Mostly single patient pre-post designs, only 1 study compared 2 cohorts with and without screening	Many studies had no f/u of negative results or reports of interval cancers Many were unclear in how patients were selected for screening Recall for additional imaging not consistently reported	2 good-quality studies (United States, Italy) with consistent cancer detection rates.	Only good-quality of screening population from Italy	2 good, 8 fair Overall: fair	From 2 good-quality studies for invasive CDR: 3.9 to 4.1 per 1,000 exams Biopsy rates: 5.9% (Italy)
	ABUS k=3	Test performance; cohort	Loss to followup only reported in 2 studies (12-20%)	Limited, only 3 fair-quality studies	Limited by unclear reporting on study population and design	3 fair Overall: Insufficient	Invasive CDR: 1.8 – 12.3 per 1,000 women/exams Recall rates: 1.5 – 13.5%
	MRI k=3	Test performance	Higher risk women, small studies	Good-quality studies Small numbers, variable findings	2 European studies, 1 from United States, all focused on high risk women	3 good but small studies Overall: Fair	All breast CDR: 3.5 to 26.8 per 1,000 exams Invasive CDR: 18 to 19 per 1,000 women/exams Recall rates: 8.6 – 23.4%
	DBT k=4	Cohorts with comparison group, Single patient pre-post	No f/u of negative tests, unclear study details with regard to invasive cancer detection, recall and biopsy rates	4 fair-quality studies	Screening populations; 3 from the United States and 1 from Italy	4 fair Overall: Fair	All breast cancer detection: 4.7 to 6.9 per 1,000 exams Recall rates: 6.6 – 10.8%
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?	k=1	RCT	Examined only effect of notification without recommendation for supplemental testing	N/A	Canadian women in organized screening program	1 good Overall: Insufficient	No measurable psychological harms from breast density notification Other harms: General risk of nephrogenic systemic sclerosis from use of gadolinium contrast in MRI for those with CKD 4-5 DBT as currently performed results in >2 times radiation as for standard digital mammogram

Abbreviations: ABUS=automated whole breast ultrasound; CDR=cancer detection rate; DBT=digital breast ultrasound; HHUS=hand-held ultrasound.

Key Question Literature Search Strategies

KQ 1: BI-RADS Accuracy and Reliability

Database: Ovid MEDLINE(R)

Search Strategy:

-
1. "breast densit*".ti,ab.
 2. parenchym*.ti,ab.
 3. mammo* pattern.ti,ab.
 4. mammo* patterns.ti,ab.
 5. radiological pattern*.ti,ab.
 6. wolfe*.ti,ab.
 7. tabar*.ti,ab.
 8. mammo* feature*.ti,ab.
 9. breast pattern*.ti,ab.
 10. mammo* densit*.ti,ab.
 11. tissue densit*.ti,ab.
 12. exp "Predictive Value of Tests"/ or diagnostic accuracy.mp. or exp "Sensitivity and Specificity"/
 13. diagnostic errors/ or observer variation/
 14. "Reproducibility of Results"/ or inter-rater.mp.
 15. or/12-14
 16. or/1-11
 17. (BI-RADS or birad or bi-rad or bi-rads or "Breast Imaging Reporting and Data System").ti,ab.
 18. 15 and 16 and 17
 19. limit 18 to ((abstracts or english language) and yr="2000 - 2014")

KQ 2, 3: Supplemental Screening Performance and Outcomes

Database: Cochrane

Search Strategy:

'mammogra AND screen* AND (breast density OR dense breast OR parenchym*) in Title, Abstract, Keywords*

Database: Ovid MEDLINE(R)

Search Strategy:

-
1. "breast densit*".ti,ab.
 2. parenchym*.ti,ab.
 3. mammo* pattern.ti,ab.
 4. mammo* patterns.ti,ab.
 5. radiological pattern*.ti,ab.
 6. wolfe*.ti,ab.

Appendix A. Detailed Methods

7. tabar*.ti,ab.
8. mammo* feature*.ti,ab.
9. breast pattern*.ti,ab.
10. mammo* densit*.ti,ab.
11. tissue densit*.ti,ab.
12. or/1-11
13. (negative test result* or false negative).mp. or exp False Negative Reactions/
14. "sensitivity and specificity"/ or "limit of detection"/ or roc curve/ or signal-to-noise ratio/
15. "sensitivity and specificity"/ or "limit of detection"/ or roc curve/ or signal-to-noise ratio/
16. or/13-15
17. ((negative adj4 mammogra*) or negative screen).mp.
18. 16 or 17
19. (supplementa* adj3 screen*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
20. (breast or mammogra*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
21. 12 and 16 and 18
22. 20 and 21
23. 12 and 19
24. (((supplementa* adj5 ultraso*) or supplementa*) adj5 imag*).mp.
25. 12 and 24
26. 20 and 25
27. 22 or 26
28. 23 or 27
29. limit 28 to ((abstracts or english language) and yr="2000 -Current")

Database: Ovid MEDLINE(R)

Search Strategy:

-
1. exp "Sensitivity and Specificity"/
 2. sensitivity.tw.
 3. specificity.tw.
 4. ((pre-test or pretest) adj probability).tw.
 5. post-test probability.tw.
 6. post-test probability.tw.
 7. likelihood ratio\$.tw.
 8. or/1-7
 9. Breast Neoplasms/
 10. (breast adj (neoplasm or neoplasms or tumour or tumor or tumors or tumours or cancer or carcinoma or carcinomas or oncologic or oncology)).mp.
 11. 9 or 10
 12. exp Mammography/
 13. Mammograph\$.ti,ab.
 14. 12 or 13

Appendix A. Detailed Methods

15. 8 and 14
16. "breast densit*".ti,ab.
17. parenchym*.ti,ab.
18. mammo* pattern.ti,ab.
19. mammo* patterns.ti,ab.
20. radiological pattern*.ti,ab.
21. wolfe*.ti,ab.
22. tabar*.ti,ab.
23. (birad* or bi-rad*).ti,ab.
24. mammo* feature*.ti,ab.
25. breast pattern*.ti,ab.
26. mammo* densit*.ti,ab.
27. tissue densit*.ti,ab.
28. "breast imaging reporting and data system".ti,ab.
29. or/16-28
30. 8 and 11 and 14 and 29
31. limit 30 to english language
1. 65. Image Processing, Computer-Assisted/ or Radiographic Image Interpretation, Computer-Assisted/ or Tomography, X-Ray Computed/ or Radiographic Image Enhancement/ or Tomography, X-Ray/ or tomosynthesis.mp. or Imaging, Three-Dimensional/
2. 66. 64 and 65
3. 67. Ultrasonography, Mammary/ or automated ultrasound.mp.
4. 68. whole breast ultrasound.mp.
5. 69. hand help ultrasound.mp.
6. 70. magnetic resonance imaging.mp. or Magnetic Resonance Imaging/
7. 71. mri.mp.
8. 72. Technetium Tc 99m Sestamibi/ or scintimammography.mp.
9. 73. or/67-72
10. 74. 31 and 73
11. 75. limit 74 to (english language and yr="2000 -Current")
12. 79. or/76-78
13. 80. 62 and 79
14. 81. limit 80 to (english language and yr="2000 -Current")
15. 82. 81 not 75
16. 83. 65 or 73
17. 84. 82 and 83

Database: Embase

Search Strategy:

-
1. 'mammography'/exp OR 'mammography' OR 'mammography system'/exp OR 'mammography system' OR mammograph*:ab,ti AND [2000-2014]/py
 2. 'dosimetry'/exp OR 'dosimetry' OR 'radiation protection'/exp OR 'radiation protection' OR 'radiation measurement'/exp OR 'radiation measurement' AND [2000-2014]/py

Appendix A. Detailed Methods

4. 'radiation exposure'/exp OR 'radiation exposure' OR 'radiation induced neoplasm'/exp OR 'radiation induced neoplasm' OR 'radiation injury'/exp OR 'radiation injury' AND [2000-2014]/py
5. 'morbidity'/exp OR 'morbidity' OR 'mortality'/exp OR 'mortality' OR 'adverse effect':ab,ti OR 'adverse effects':ab,ti OR harm:ab,ti OR harms:ab,ti OR contraindic*:ab,ti AND [2000-2014]/py
6. #2 OR #4
7. #1 AND #5 AND #6
- 8.1 'breast tumor'/exp/dm_pc,dm_di
- 8.2 (breast NEXT/5 (neoplasm* OR tumour* OR tumor* OR cancer* OR carcinom* OR oncolog*)):ab,ti
- 8.3 #8.1 OR #8.2
- 8.4 'mass screening'/exp OR 'mass radiography'/exp
- 8.5 'neoplasm'/exp/dm_pc,dm_di
- 8.6 'mammography'/exp OR 'mammography system'/exp OR mammograph*:ab,ti
- 8.7 screen*:ab,ti
- 8.8 #8.4 OR #8.5 OR #8.6 OR #8.7
- 8.9 #8.3 AND #8.8
- 8.10 'sensitivity and specificity'/exp OR sensitivity:ab,ti OR specificity:ab,ti
- 8.11 (('pre test' OR pretest) NEAR/5 probability):ab,ti
- 8.12 (('pre test' OR pretest) NEAR/5 probability):ab,ti
- 8.13 'likelihood ratio':ab,ti OR 'likelihood ratios':ab,ti
- 8.14 #8.10 OR #8.11 OR #8.12 OR #8.13
- 8.15 #8.9 AND #8.14
- 8.16 'breast density':ab,ti OR 'dense breasts':ab,ti OR 'dense breast':ab,ti OR parenchym*:ab,ti OR 'mammographic feature':ab,ti OR 'mammographic features':ab,ti OR (mammography NEAR/2 feature*):ab,ti OR 'breast pattern':ab,ti OR 'breast patterns':ab,ti OR (breast NEAR/3 pattern):ab,ti OR 'mammographic density':ab,ti OR (mammography NEAR/3 density):ab,ti OR 'mammographic pattern':ab,ti OR 'mammographic patterns':ab,ti OR (mammography NEAR/2 patterns):ab,ti OR 'radiological pattern':ab,ti OR 'radiological patterns':ab,ti OR wolfe*:ab,ti OR tabar*:ab,ti OR birad*:ab,ti OR 'bi rad':ab,ti OR 'breast imaging reporting and data system':ab,ti OR 'tissue density':ab,ti OR (tissue NEAR/3 density):ab,ti
- 8.17 #8.15 AND #8.16
- 8.18 #8.17 AND [english]/lim AND [2000-2014]/py

KQ 4: Breast Density Notification-Related Harms

Database: Ovid MEDLINE(R)

Search Strategy:

-
1. "breast densit*" .ti,ab.
 2. parenchym* .ti,ab.
 3. mammo* pattern.ti,ab.
 4. mammo* patterns.ti,ab.
 5. radiological pattern* .ti,ab.
 6. wolfe* .ti,ab.

Appendix A. Detailed Methods

7. tabar*.ti,ab.
8. mammo* feature*.ti,ab.
9. breast pattern*.ti,ab.
10. mammo* densit*.ti,ab.
11. tissue densit*.ti,ab.
12. or/1-11
13. px.fs.
14. "Risk Factors"/
15. "Health Knowledge, Attitudes, Practice"/
16. "Awareness"/
17. or/13-16
18. 12 and 17
19. (breast or mammogr*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
20. 18 and 19
21. "False Positive Reactions"/ or "False Negative Reactions"/ or "Anxiety"/ or "Depression"/ or "Stress, Psychological"/ or "Patient Acceptance of Health Care"/ or px.fs. or "Motivation"/ or barrier\$.mp. or "attitude to health"/ or "womens health"/ or exp "Quality of Life"/ or exp Health Status Indicators/
22. 20 and 21
23. limit 22 to ((abstracts or english language) and yr="2000 - current")

Database: Embase

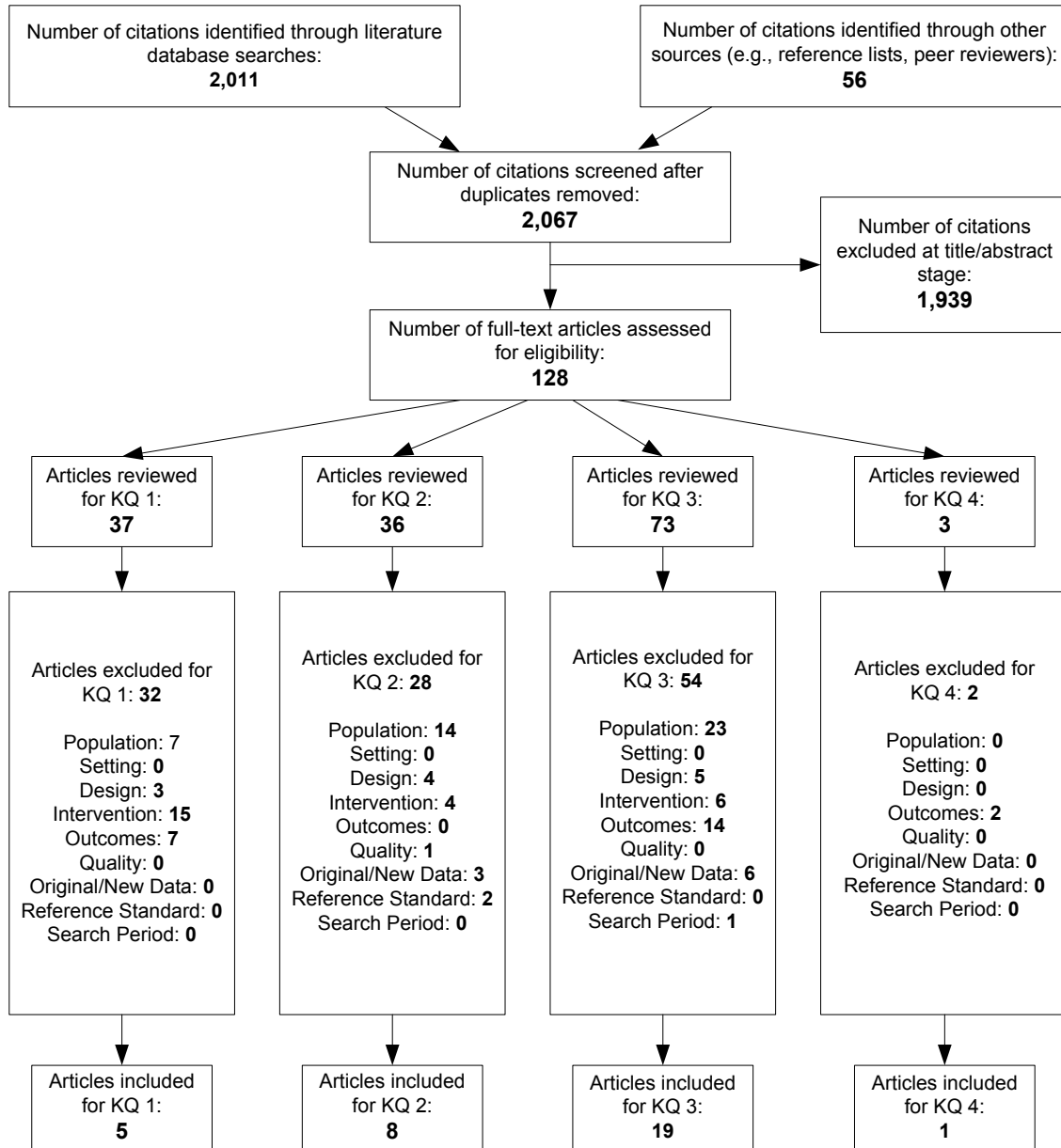
Search Strategy:

-
1. 'breast density':ab,ti OR (breast NEAR/3 density):ab,ti OR 'dense breast':ab,ti OR 'dense breasts':ab,ti
 2. parenchym*:ab,ti
 3. (mammo* NEAR/1 pattern):ab,ti
 4. (radiologic* NEAR/1 pattern):ab,ti
 5. #breastcascreeing-59 OR #breastcascreeing-58 OR #breastcascreeing-57 OR #breastcascreeing-56 OR #breastcascreeing-55 OR #breastcascreeing-54 OR #breastcascreeing-53
 6. #1 OR #2 OR #3 OR #4 OR #5
 7. 'false positive reactions'/exp OR 'false positive reactions' OR 'false negative reactions'/exp OR 'false negative reactions' OR 'anxiety'/exp OR 'anxiety' OR 'depression'/exp OR 'depression' OR 'stress, psychological'/exp OR 'stress, psychological' OR 'patient acceptance of health care'/exp OR 'patient acceptance of health care' OR psychological OR 'motivation'/exp OR 'motivation' OR barrier\$:ab,ti OR 'attitude to health'/exp OR 'attitude to health' OR 'womens health'/exp OR 'womens health' OR 'quality of life'/exp OR 'quality of life' OR 'health'/exp OR health AND status AND indicators OR harms OR harm AND ('reduction'/exp OR reduction) OR 'risk'/exp OR risk AND assessment
 8. #6 AND #7
 9. mammogram OR 'mammography'/exp OR mammography OR magnetic AND resonance AND ('imaging'/exp OR imaging) OR 'mri'/exp OR mri OR ultrasonog*

Appendix A. Detailed Methods

10. #8 AND #9
11. 'breast'/exp OR breast
12. #10 AND #11
13. #12 AND [english]/lim AND [2000-2014]/py
14. harms AND [2000-2014]/py
15. #13 AND #14
17. supplementary AND ('screening'/exp OR screening) AND [2000-2014]/py
18. #13 AND #17
19. #15 OR #18
20. 'risk assessment'/exp
22. #6 AND #20 AND [2000-2014]/py
23. #11 AND #22
24. #1 AND #23
25. #24 AND [english]/lim AND [2000-2014]/py

Appendix A Figure 1. Literature Flow Diagram



Appendix A Table 1. Inclusion and Exclusion Criteria

Category	KQs	Inclusion	Exclusion
Populations	1	Women primarily aged 40 years and older receiving screening mammography (digital or film)	Women with: <ul style="list-style-type: none"> • Pre-existing breast cancer • Clinically significant BRCA 1/2 mutations • Li-Fraumeni syndrome • Cowden syndrome • Hereditary diffuse gastric syndrome • Other familial breast cancer syndromes • High-risk breast lesions (DCIS, LCIS, ADH, ALH) • Previous doses of chest radiation (>20Gy) before age 30 • Undergoing diagnostic or surveillance mammography
	2-4	Women primarily aged 40 years and older undergoing screening mammography or who had a negative mammogram and found to have dense breasts. Dense breasts defined as BI-RADS 3 or 4, c or d, or “heterogeneous” or “extremely” dense	
Setting	1-4	Conducted in primary care or other setting with primary care-comparable population	Settings not generalizable to primary care
Intervention or Exposure	1	Breast Imaging Reporting and Data System (BI-RADS) for mammographic breast density	Digital or full-film mammography alone (okay for KQ1); use of new technologies for diagnostic or surveillance purposes; use in a diagnostic or surveillance setting only
	2-3	Breast MRI, hand-held ultrasound, whole breast ultrasound, and/or breast DBT used in a screening setting	
	4	Received breast density notification	
Comparisons or Nonexposure	1	Other approaches for breast density determination (BIRAD ultrasound, percent density, other systems (e.g., Boyd, Wolfe):	
	2-3	Digital mammography; women ages 40-49 vs. 50-59 vs. 60-69 vs. 70-79 (or other age comparisons); non-dense breasts	
	4	Did not receive breast density notification	
Outcomes	1	BI-RADS density determination concordance (inter- and intra-rater reliability, consistency)	
	2	Test performance characteristics (sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios for invasive breast cancers, breast lesions [DCIS], total breast cancers, breast cancers by stage)	
	3	Health Outcomes (interval cancer incidence, DCIS diagnosis rate, stage at breast cancer diagnosis, invasive breast cancer recurrence rate, breast cancer mortality)	
	4	Harms of breast density notification (quality of life, anxiety)	
Study Designs	1	RCT’s, prospective and retrospective cohort studies, cross-sectional studies, test sets involving multiple blinded readings of digital or film mammograms by at least three readers looking at reproducibility cross-sectional or longitudinally	Test sets involving fewer than 3 independent readers
	2	Diagnostic accuracy studies with reference standard and more than one radiologist/reader, cohort studies with more than one radiologist/reader, and meta-analyses	Narrative reviews, editorials, commentary
	3	Randomized controlled trials (RCTs), cohort studies ; meta-analyses	

Appendix A Table 1. Inclusion and Exclusion Criteria

Category	KQs	Inclusion	Exclusion
	4	RCTs, prospective and retrospective cohort studies, case-control studies, cross-sectional studies, systematic reviews, meta-analyses, and modeling studies	
Language	1-4	English only	Non-English languages
Publication Date	1-4	Trials published from January 2000 to present	Trials published before January 2000
Study Quality	1-4	Fair- and good-quality studies	Poor-quality studies

Appendix B. Ongoing Studies and Trials Pending Assessment

Investigator (Location) Study Title/Name	Number of Participants/ Estimated Enrollment	Intervention	Outcomes	2015 Status
Carla van Gils (Netherlands) <i>Breast Cancer Screening with MRI in Women Aged 50-75 Years with Extremely Dense Breast Tissue: the DENSE Trial</i>	36,185	Biennial screening with mammography and MRI compared to mammography alone among women with >75% mammographic density	Interval cancers; screen-detected tumors, tumor size, stage, grade distribution; referral rate; positive predictive value and false positives; biopsies per positive test; mortality rate; cost effectiveness; quality of life	Study Period: November 2011 – December 2019 Recruiting
Alberto Tagliafico (Italy) <i>Tomosynthesis (TS) Versus Ultrasonography (US) in Women with Dense Breasts (ASTOUND)</i>	4,000	Screening with DBT compared to ultrasound in women with dense breasts	Sensitivity; specificity	Study Period: December 2012 – July 2016 Recruiting
Denise Chough (Pittsburg) <i>Assessment of Automated Breast Ultrasound</i>	1,000	Screening with ABUS, DBT, or a combination of both	False positive rates; biopsy rates; NPV	Study Period: April 2015 – November 2017 Not yet recruiting
Constance Lehman (Seattle) <i>Automated Breast Ultrasound and Digital Breast Tomosynthesis Screening Compared to Full Field Digital Mammography in Women With Dense Breasts</i>	650	Screening with digital mammography and DBT with and without automated whole breast ultrasound in women with dense breasts	Abnormal interpretation rate; sensitivity; specificity; cancer rate; positive predictive value; negative predictive value	Study Period: April 2014 – June 2017 Recruiting
Jung Min Chang (South Korea) <i>Comparison of Diagnostic Performance of Digital Breast Tomosynthesis and Ultrasound in Women With Dense Breasts</i>	825	Screening with DBT compared to ultrasonography in women with dense breasts	Sensitivity; specificity; negative predictive value; positive predictive value	Study Period: June 2013 – April 2014 Recruiting
Tilanus-Linthorst (Netherlands) <i>Breast density in women with familial risk as indicator for the use of mammography or MRI to screen for breast cancer: An RCT (FaMRisc)</i>	2,000	Annual screening with MRI and mammography versus mammography alone in women at high risk of breast cancer	Cancer detection; interval cancer rates; stage at diagnosis; sensitivity; specificity; mortality; cost-effectiveness	Study Period: January 2011 – January 2015 Recruiting

Appendix C. Excluded Studies

Exclusion Code

- E1. Population
 - a. Other definition of dense breasts
- E2. Setting
- E3. Intervention or Exposure
 - a. Not an included modality
 - b. Diagnostic or surveillance use
 - c. Did not utilize BI-RADS assessment and compare to other approaches
- E4. Study design
 - a. Inadequate number of readers
- E5. No relevant outcomes
- E6. Study quality
- E7. No original data to include; publication or dataset with longer followup, more complete data, or same data already included in review
- E8. No reference standard utilized
- E9. Precedes search period

1. Abdolell M, Tsuruda K, Schaller G, Caines J. Statistical evaluation of a fully automated mammographic breast density algorithm. *Comput Math Methods Med*. 2013;2013:651091. PMID: 23737861. **KQ3E5**
2. Bakic PR, Carton AK, Kontos D, Zhang C, Troxel AB, Maidment AD. Breast percent density: estimation on digital mammograms and central tomosynthesis projections. *Radiology*. 2009;252(1):40-9. PMID: 19420321. **KQ1E3c**
3. Benndorf M, Baltzer PA, Vag T, Gajda M, Runnebaum IB, Kaiser WA. Breast MRI as an adjunct to mammography: Does it really suffer from low specificity? A retrospective analysis stratified by mammographic BI-RADS classes. *Acta Radiol*. 2010;51(7):715-21. PMID: 20707656. **KQ2E3b**
4. Benson SR, Blue J, Judd K, Harman JE. Ultrasound is now better than mammography for the detection of invasive breast cancer. *Am J Surg*. 2004;188(4):381-5. PMID: 15474430. **KQ3E1**
5. Berg WA. Supplemental screening sonography in dense breasts. *Radiol Clin North Am*. 2004;42(5):845-51, vi. PMID: 15337420. **KQ3E4**
6. Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Bohm-Velez M, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. *JAMA - Journal of the American Medical Association*. 2008;299(18):2151-63. PMID: 18477782. **KQ2E7,KQ3E7**
7. Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR American Journal of Roentgenology*. 2000;174(6):1769-77. PMID: 10845521. **KQ1E5**
8. Blane CE, Fitzgerald JT, Gruppen LD, Oh MS, Helvie MA, Andersson I. Decreasing rate of fatty involution at screening mammography. *Acad Radiol*. 2002; 9(8):895-8. PMID: 12186437. **KQ3E5**
9. Boudreau DM, Rutter CM, Buist DS. The influence of statin use on breast density. *Cancer Epidemiol Biomarkers Prev*. 2006;15(5):1026-9. PMID: 16702387. **KQ3E5**
10. Brem RF, Hoffmeister JW, Rapelyea JA, Zisman G, Mohtashemi K, Jindal G, et al. Impact of breast density on computer-aided detection for breast cancer. *American Journal of Roentgenology*. 2005;184(2):439-44. PMID: 15671360. **KQ2E3a, KQ3E3a**
11. Buchberger W, Niehoff A, Obrist P, DeKoekkoek-Doll P, Dunser M. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Semin Ultrasound CT MR*. 2000;21(4):325-36. PMID: 11014255. **KQ2E1, KQ3E1**
12. Chae EY, Kim HH, Cha JH, Shin HJ, Kim H. Evaluation of screening whole-breast sonography as a supplemental tool in conjunction with mammography in women with dense breasts. *J Ultrasound Med*. 2013;32(9):1573-8. PMID: 23980217. **KQ2E3, KQ3E3**

Appendix C. Excluded Studies

13. Chang RF, Chang-Chien KC, Takada E, Suri JS, Moon WK, Wu JH, et al. Three comparative approaches for breast density estimation in digital and screen film mammograms. Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine & Biology Society. 2006;1:4853-6. PMID: 17945860. **KQ1E3c**
14. Ciatto S, Bernardi D, Calabrese M, Durando M, Gentilini MA, Mariscotti G, et al. A first evaluation of breast radiological density assessment by QUANTRA software as compared to visual classification. *Breast*. 2012;21(4):503-6. PMID: 22285387. **KQ1E5**
15. Ciatto S, Houssami N, Apruzzese A, Bassetti E, Brancato B, Carozzi F, et al. Categorizing breast mammographic density: intra- and interobserver reproducibility of BI-RADS density categories. *Breast*. 2005;14(4):269-75. PMID: 16085233. **KQ1E1**
16. Cohen SL, Margolies LR, Schwager SJ, Zuckerman S, Patel N, Szabo J, et al. Early discussion of breast density and supplemental breast cancer screening: is it possible? *Breast Journal*. 2014;20(3):229-34. PMID: 24689788. **KQ4E5**
17. Corsetti V, Ferrari A, Ghirardi M, Bergonzini R, Bellarosa S, Angelini O, et al. Role of ultrasonography in detecting mammographically occult breast carcinoma in women with dense breasts. *Radiologia Medica*. 2006;111(3):440-8. PMID: 16683089. **KQ3E7**
18. Corsetti V, Houssami N, Ferrari A, Ghirardi M, Bellarosa S, Angelini O, et al. Breast screening with ultrasound in women with mammography-negative dense breasts: evidence on incremental cancer detection and false positives, and associated cost (Structured abstract). *Eur J Cancer*. 2008;44(4):539-44. PMID: 18267357. **KQ3E7**
19. Crystal P, Strano SD, Shcharynski S, Koretz MJ. Using sonography to screen women with mammographically dense breasts. *AJR Am J Roentgenol*. 2003;181(1):177-82. PMID: 12818853. **KQ2E1a**
20. De Felice C, Savelli S, Angeletti M, Ballesio L, Manganaro L, Meggiorini ML, et al. Diagnostic utility of combined ultrasonography and mammography in the evaluation of women with mammographically dense breasts. *J Ultrasound*. 2007;10(3):143-51. PMID: 23396266. **KQ2E4a**
21. de Gelder R, van As E, Tilanus-Linthorst MM, Bartels CC, Boer R, Draisma G, et al. Breast cancer screening: evidence for false reassurance? *Int J Cancer*. 2008;123(3):680-6. PMID: 18484587. **KQ3E3b**
22. Del Carmen MG, Hughes KS, Halpern E, Rafferty E, Kopans D, Parisky YR, et al. Racial differences in mammographic breast density. *Cancer*. 2003;98(3):590-6. PMID: 17377060. **KQ3E5**
23. DeMartini WB, Liu F, Peacock S, Eby PR, Gutierrez RL, Lehman CD. Background parenchymal enhancement on breast MRI: impact on diagnostic performance. *AJR American Journal of Roentgenology*. 2012;198(4):W373-80. PMID: 22451576. **KQ2E1, KQ3E1**
24. Ding J, Warren R, Warsi I, Day N, Thompson D, Brady M, et al. Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: case-control study. *Cancer Epidemiol Biomarkers Prev*. 2008;17(5):1074-81. PMID: 18483328. **KQ3E5**
25. Drukker K, Horsch KJ, Pesce LL, Giger ML. Interreader scoring variability in an observer study using dual-modality imaging for breast cancer detection in women with dense breasts. *Acad Radiol*. 2013;20(7):847-53. PMID: 23601952. **KQ3E3**
26. Elmore JG, Miglioretti DL, Reisch LM, Barton MB, Kreuter W, Christiansen CL, et al. Screening mammograms by community radiologists: variability in false-positive rates. *J Natl Cancer Inst*. 2002;94(18):1373-80. PMID: 12237283. **KQ3E5**
27. Fowler EE, Sellers TA, Lu B, Heine JJ. Breast Imaging Reporting and Data System (BI-RADS) breast composition descriptors: Automated measurement development for full field digital mammography. *Med Phys*. 2013;40(11). PMID: 24320473. **KQ1E3c**
28. Gao J, Warren R, Warren-Forward H, Forbes JF. Reproducibility of visual assessment on mammographic density. *Breast Cancer Res Treat*. 2008;108(1):121-7. PMID: 17616811. **KQ1E3c**
29. Garrido-Esteba M, Ruiz-Perales F, Miranda J, Asuncion N, Gonzalez-Roman I, Sanchez-Contador C, et al. Evaluation of mammographic density patterns: reproducibility and concordance among scales. *BMC Cancer*. 2010;10:485. PMID: 20836850. **KQ2E4a**

Appendix C. Excluded Studies

30. Gartlehner G, Thaler K, Chapman A, Kaminski-Hartenthaler A, Berzaczy D, Van Noord Megan G, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. *Cochrane Database of Syst Rev.* 2013;4. DOI: 10.1002/14651858.CD009632.pub2. **KQ2E1, KQ3E1**
31. Gartlehner G, Thaler KJ, Chapman A, Kaminski A, Berzaczy D, Van Noord MG, et al. Adjunct ultrasonography for breast cancer screening in women at average risk: a systematic review. *International Journal of Evidence-Based Healthcare.* 2013;11(2):87-93. PMID: 23750571. **KQ2E7, KQ3E7**
32. Granader EJ, Dwamena B, Carlos RC. MRI and mammography surveillance of women at increased risk for breast cancer: recommendations using an evidence-based approach. *Acad Radiol.* 2008;15(12):1590-5. PMID: 19000876. **KQ3E1**
33. 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(3):694-700. PMID: 23901124. **KQ2E8**
34. Health Quality O. Cancer screening with digital mammography for women at average risk for breast cancer, magnetic resonance imaging (MRI) for women at high risk: an evidence-based analysis. *Ontario Health Technology Assessment Series.* 2010; 10(3):1-55. PMID: 23074406. **KQ2E1, KQ3E1**
35. Heine JJF, E. E. E.; Flowers, C. I. Full Field Digital Mammography and Breast Density. Comparison of Calibrated and Noncalibrated Measurements. *Acad Radiol.* 2011;18(11):1430-6. PMID: 21971260. **KQ1E3c**
36. Hollingsworth AB, Stough RG. An alternative approach to selecting patients for high-risk screening with breast MRI. *Breast Journal.* 2014;20(2):192-7. PMID: 24387050. **KQ3E4**
37. Honjo S, Ando J, Tsukioka T, Morikubo H, Ichimura M, Sunagawa M, et al. Relative and combined performance of mammography and ultrasonography for breast cancer screening in the general population: a pilot study in Tochigi Prefecture, Japan. *Jpn J Clin Oncol.* 2007;37(9):715-20. PMID: 17766996. **KQ3E1**
38. Irwig L, Houssami N, van Vliet C. New technologies in screening for breast cancer: a systematic review of their accuracy. *Br J Cancer.* 2004;90(11):2118-22. PMID: 15150556. **KQ3E1**
39. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. *Radiology.* 2001;221(3):641-9. PMID: 11719658. **KQ2E4a, KQ3E4a**
40. Keller BM, Nathan DL, Gavenonis SC, Chen J, Conant EF, Kontos D. Reader variability in breast density estimation from full-field digital mammograms: the effect of image postprocessing on relative and absolute measures. *Acad Radiol.* 2013;20(5):560-8. PMID: 23465381. **KQ1E3c**
41. Kerlikowske K, Grady D, Barclay J, Frankel SD, Ominsky SH, Sickles EA, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst.* 1998;90(23):1801-9. PMID: 9839520. **KQ1E4**
42. Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. *Ann Intern Med.* 2011;155(8):493-502. PMID: 22007043. **KQ3E4a**
43. Kim do Y, Moon WK, Cho N, Ko ES, Yang SK, Park JS, et al. MRI of the breast for the detection and assessment of the size of ductal carcinoma in situ. *Korean Journal of Radiology.* 2007;8(1):32-9. PMID: 17277561. **KQ3E3b**
44. Kim SJ, Chang JM, Cho N, Chung SY, Han W, Moon WK. Outcome of breast lesions detected at screening ultrasonography. *Eur J Radiol.* 2012;81(11):3229-33. PMID: 22591758. **KQ3E5**
45. Kim WH, Moon WK, Kim SJ, Yi A, Yun BL, Cho N, et al. Ultrasonographic assessment of breast density. *Breast Cancer Res Treat.* 2013;138(3):851-9. PMID: 23568482. **KQ1E1**
46. Kim WH, Moon WK, Kim SM, Yi A, Chang JM, Koo HR, et al. Variability of breast density assessment in short-term reimaging with digital mammography. *Eur J Radiol.* 2013;82(10):1724-30. PMID: 23727379. **KQ1E1**
47. Ko SY, Kim EK, Kim MJ, Moon HJ. Mammographic density estimation with automated volumetric breast density measurement. *Korean Journal of Radiology.* 2014;15(3):313-21. PMID: 24843235. **KQ1E3c**
48. Kolb TM, Lichy J, Newhouse JH. Occult cancer in women with dense breasts: detection with screening US--diagnostic yield and tumor characteristics. *Radiology.* 1998;207(1):191-9. PMID: 9530316. **KQ3E9**

Appendix C. Excluded Studies

49. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology*. 2002;225(1):165-75. PMID: 12355001. **KQ2E6**
50. Kriege M, Brekelmans CT, Boetes C, Besnard PE, Zonderland HM, Obdeijn IM, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *N Engl J Med*. 2004;351(5):427-37. PMID: 15282350. **KQ3E1**
51. Kriege M, Brekelmans CT, Peterse H, Obdeijn IM, Boetes C, Zonderland HM, et al. Tumor characteristics and detection method in the MRISC screening program for the early detection of hereditary breast cancer. *Breast Cancer Res Treat*. 2007;102(3):357-63. PMID: 17051427. **KQ3E1**
52. Kuhl CK, Schrading S, Bieling HB, Wardelmann E, Leutner CC, Koenig R, et al. MRI for diagnosis of pure ductal carcinoma in situ: a prospective observational study. *Lancet*. 2007;370(9586):485-92. PMID: 17693177. **KQ2E1**
53. Leconte I, Feger C, Galant C, Berliere M, Berg BV, D'Hoore W, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. *AJR American Journal of Roentgenology*. 2003;180(6):1675-9. PMID: 12760942. **KQ2E8**
54. Lee CI, Ichikawa L, Rochelle MC, Kerlikowske K, Miglioretti DL, Sprague BL, et al. Breast MRI BI-RADS Assessments and Abnormal Interpretation Rates by Clinical Indication in US Community Practices. *Acad Radiol*. 2014. PMID: 25126973. **KQ1E3c**
55. Lehman CD, Blume JD, Weatherall P, Thickman D, Hylton N, Warner E, et al. Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. *Cancer*. 2005;103(9):1898-905. PMID: 15800894. **KQ2E1**
56. Leung TK, Huang PJ, Chen CS, Lin YH, Wu CH, Lee CM. Is breast MRI screening more effective than digital mammography in Asian women? *Journal of Experimental and Clinical Medicine*. 2010;2(5):245-50. **KQ3E5**
57. Li J, Szekely L, Eriksson L, Heddsom B, Sundbom A, Czene K, et al. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. *Breast Cancer Res*. 2012;14(4). PMID: 22846386. **KQ3E5**
58. Lord SJ, Lei W, Craft P, Cawson JN, Morris I, Walleser S, et al. A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer. *Eur J Cancer*. 2007;43(13):1905-17. PMID: 17681781. **KQ2E1, KQ3E1**
59. Lowry SJ, Aiello Bowles EJ, Anderson ML, Buist DSM. Predictors of breast density change after hormone therapy cessation: Results from a randomized trial. *Cancer Epidemiology Biomarkers and Prevention*. 2011;20(10):2309-12. PMID: 21813728. **KQ3E5**
60. Manning MA, Duric N, Littrup P, Bey-Knight L, Penner L, Albrecht TL. Knowledge of breast density and awareness of related breast cancer risk. *J Cancer Educ*. 2013;28(2):270-4. PMID: 23467999. **KQ4E5**
61. Mercer CE, Hogg P, Kelly J, Borgen R, Millington SR, Hilton B, et al. A mammography image set for research purposes using BI-RADS density classification. *Radiol Technol*. 2014;85(6):609-13. PMID: 25002640. **KQ1E1**
62. Myles JP, Salminen T, Duffy SW, Prevost TC, Day NE, Hakama M. Prospective estimation of rates of change in mammographic parenchymal patterns: Influence of age and of hormone replacement therapy. *Breast*. 2004;13(1):56-60. PMID: 14759717. **KQ3E5**
63. Nicholson BT, LoRusso AP, Smolkin M, Bovbjerg VE, Petroni GR, Harvey JA. Accuracy of assigned BI-RADS breast density category definitions. *Acad Radiol*. 2006;13(9):1143-9. PMID: 16935726. **KQ1E5**
64. Nickson C, Arzhaeva Y, Aitken Z, Elgindy T, Buckley M, Li M, et al. AutoDensity: An automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes. *Breast Cancer Res*. 2013;15(5). PMID: 24020331. **KQ1E3c**
65. Nothacker M, Duda V, Hahn M, Warm M, Degenhardt F, Madjar H, et al. Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review. *BMC Cancer*. 2009;9:335. PMID: 19765317. **KQ2E7, KQE7**
66. Obdeijn IM, Loo CE, Rijnsburger AJ, Wasser MN, Bergers E, Kok T, et al. Assessment of false-negative cases of breast MR imaging in women with a familial or genetic predisposition. *Breast Cancer Res Treat*. 2010;119(2):399-407. PMID: 19876732. **KQ3E4**

Appendix C. Excluded Studies

67. O'Driscoll D, Warren R, MacKay J, Britton P, Day NE. Screening with breast ultrasound in a population at moderate risk due to family history. *J Med Screen.* 2001;8(2):106-9. PMID: 11480440. **KQ3E1**
68. Ohuchi N, Suzuki A, Sakurai Y, Kawai M, Narikawa Y, Narimatsu H, et al. Current status and problems of breast cancer screening. *Japan Medical Association Journal.* 2009;52(1):45-9. **KQ3E4**
69. Oliver A, Llado X, Freixenet J, Marti R, Perez E, Pont J, et al. Influence of Using Manual or Automatic Breast Density Information in a Mass Detection CAD System. *Acad Radiol.* 2010;17(7):877-83. PMID: 20540910. **KQ1E3c**
70. Ooms EA, Zonderland HM, Eijkemans MJ, Kriege M, Mahdavian Delavary B, Burger CW, et al. Mammography: interobserver variability in breast density assessment. *Breast.* 2007;16(6):568-76. PMID: 18035541. **KQ1E1**
71. Pediconi F, Catalano C, Roselli A, Dominelli V, Cagioli S, Karatasios A, et al. The challenge of imaging dense breast parenchyma: is magnetic resonance mammography the technique of choice? A comparative study with x-ray mammography and whole-breast ultrasound. *Invest Radiol.* 2009;44(7):412-21. PMID: 19448554. **KQ2E1, KQ3E1**
72. Perez-Gomez B, Ruiz F, Martinez I, Casals M, Miranda J, Sanchez-Contador C, et al. Women's features and inter-/intra-rater agreement on mammographic density assessment in full-field digital mammograms (DDM-SPAIN). *Breast Cancer Res Treat.* 2012;132(1):287-95. PMID: 22042363. **KQ1E3c**
73. Peters G, Anderson J, Longman G, Thomson J, Taylor D, Bennett M, et al. Magnetic resonance findings in women at high risk for developing breast cancer: an Australian feasibility study. *J Med Imaging Radiat Oncol.* 2008;52(1):29-35. PMID: 18373823. **KQ3E1**
74. Podo F, Sardanelli F, Canese R, D'Agnolo G, Natali PG, Crecco M, et al. The Italian multi-centre project on evaluation of MRI and other imaging modalities in early detection of breast cancer in subjects at high genetic risk. *J Exp Clin Cancer Res.* 2002;21(3 Suppl):115-24. PMID: 12585665. **KQ3E1**
75. Riedl CC, Ponhold L, Flory D, Weber M, Kroiss R, Wagner T, et al. Magnetic resonance imaging of the breast improves detection of invasive cancer, preinvasive cancer, and premalignant lesions during surveillance of women at high risk for breast cancer. *Clin Cancer Res.* 2007;13(20):6144-52. PMID: 17947480. **KQ3E1**
76. Rubinstein WS, Latimer JJ, Sumkin JH, Huerbin M, Grant SG, Vogel VG. Prospective screening study of 0.5 Tesla dedicated magnetic resonance imaging for the detection of breast cancer in young, high-risk women. *BMC Womens Health.* 2006;6:10. PMID: 16800895. **KQ3E1**
77. Sadowski EA, Kelcz F. Frequency of malignancy in lesions classified as probably benign after dynamic contrast-enhanced breast MRI examination. *J Magn Reson Imaging.* 2005;21(5):556-64. PMID: 15834907. **KQ3E1**
78. Sardanelli F, Giuseppetti GM, Panizza P, Bazzocchi M, Fausto A, Simonetti G, et al. Sensitivity of MRI versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard. *American Journal of Roentgenology.* 2004; 183(4):1149-57. PMID: 15385322. **KQ2E1**
79. Sardanelli F, Podo F. Breast MR imaging in women at high-risk of breast cancer. Is something changing in early breast cancer detection? *Eur Radiol.* 2007;17(4):873-87. PMID: 17008989. **KQ2E4**
80. Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts. *Am J Obstet Gynecol.* 2014. PMID: 24959654. **KQ3E7**
81. Seo JM, Ko ES, Han BK, Ko EY, Shin JH, Hahn SY. Automated volumetric breast density estimation: a comparison with visual assessment. *Clin Radiol.* 2013;68(7):690-5. PMID: 23434202. **KQ1E5**
82. Sergeant JC, Walshaw L, Wilson M, Seed S, Barr N, Beetles U, et al., editors. Same task, same observers, different values: the problem with visual assessment of breast density. *Medical Imaging Conference: Image Perception, Observer Performance, and Technology Assessment; 2013 Feb 09; Lake Buena Vista, Florida: International Society for Optics and Photonics; 2013. KQ1E3c*

Appendix C. Excluded Studies

83. Singh JM, Fallenberg EM, Diekmann F, Renz DM, Witlandt R, Bick U, et al. Volumetric breast density assessment: reproducibility in serial examinations and comparison with visual assessment. *Rofo: Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin*. 2013;185(9):844-8. PMID: 23888472. **KQ1E1**
84. Sprague BL, Stout NK, Schechter C, van Ravesteyn NT, Cevik M, Oguzhan A, et al. Potential impact of legislation mandating breast density notification: benefits, harms, and cost effectiveness of supplemental ultrasound screening. *Ann Intern Med*. in press. **KQ1E5**
85. 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(10):25217577. **KQ1E5**
86. Stone J, Ding J, Warren RML, Duffy SW, Hopper JL. Using mammographic density to predict breast cancer risk: Dense area or percentage dense area. *Breast Cancer Res*. 2010;12(6). PMID: 21087468. **KQ3E5**
87. Stone J, Warren RM, Pinney E, Warwick J, Cuzick J. Determinants of percentage and area measures of mammographic density. *Am J Epidemiol*. 2009;170(12):1571-8. PMID: 19910376. **KQ1E3c**
88. Stone JG, A.; Martin, L. J.; Yaffe, M.; Minkin, S.; Boyd, N. F. The detection of change in mammographic density. *Cancer Epidemiol Biomarkers Prev*. 2003; 12(7):625-30. PMID: 12869401. **KQ3E5**
89. Stoutjesdijk MJ, Boetes C, Jager GJ, Beex L, Bult P, Hendriks JH, et al. Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. *J Natl Cancer Inst*. 2001; 93(14):1095-102. PMID: 11459871. **KQ2E1, KQ3E1**
90. Svahn TM, Chakraborty DP, Ikeda D, Zackrisson S, Do Y, Mattsson S, et al. Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. *Br J Radiol*. 2012;85(1019):e1074-82. PMID: 22674710. **KQ3E1**
91. Tagliafico A, Tagliafico G, Astengo D, Cavagnetto F, Rosasco R, Rescinito G, et al. Mammographic density estimation: one-to-one comparison of digital mammography and digital breast tomosynthesis using fully automated software. *Eur Radiol*. 2012;22(6):1265-70. PMID: 22358426. **KQ1E4**
92. Tagliafico A, Tagliafico G, Tosto S, Chiesa F, Martinoli C, Derchi LE, et al. Mammographic density estimation: comparison among BI-RADS categories, a semi-automated software and a fully automated one. *Breast*. 2009;18(1):35-40. PMID: 19010678. **KQ1E5**
93. Tagliafico AS, Tagliafico G, Cavagnetto F, Calabrese M, Houssami N. Estimation of percentage breast tissue density: Comparison between digital mammography (2D full field digital mammography) and digital breast tomosynthesis according to different BI-RADS categories. *Br J Radiol*. 2013;86(1031). PMID: 24029631. **KQ1E1**
94. Tilanus-Linthorst MMA, Obdeijn IMM, Bartels KCM, De Koning HJ, Oudkerk M. First experiences screening women at high risk for breast cancer with MR imaging. *Breast Cancer Res Treat*. 2000;63(1):53-60. PMID: 11079159. **KQ2E1**
95. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: High-resolution x-ray imaging observer study. *Radiology*. 2012;262(3):788-96. PMID: 22274840. **KQ2E1a**
96. Wang FL, Chen F, Yin H, Xu N, Wu XX, Ma JJ, et al. Effects of age, breast density and volume on breast cancer diagnosis: a retrospective comparison of sensitivity of mammography and ultrasonography in China's rural areas. *Asian Pacific Journal of Cancer Prevention: Apjcp*. 2013;14(4):2277-82. PMID: 23725127. **KQ2E1**
97. Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. *Ann Intern Med*. 2008;148(9):671-9. PMID: 18458280. **KQ2E1, KQ3E1**
98. Warren RM, Pointon L, Caines R, Hayes C, Thompson D, Leach MO, et al. What is the recall rate of breast MRI when used for screening asymptomatic women at high risk? *Magn Reson Imaging*. 2002;20(7):557-65. PMID: 12413602. **KQ3E1**
99. Warren RM, Young JR, McLean L, Lyons K, Wilson AR, Evans A, et al. Radiology review of the UKCCCR Breast Screening Frequency Trial: potential improvements in sensitivity and lead time of radiological signs. *Clin Radiol*. 2003;58(2):128-32. PMID: 1262304. **KQ2E3**

Appendix C. Excluded Studies

100. Weinstein SP, Localio AR, Conant EF, Rosen M, Thomas KM, Schnall MD. Multimodality screening of high-risk women: a prospective cohort study. *J Clin Oncol*. 2009;27(36):6124-8. PMID: 19884532. **KQ3E1**

101. Wojcinski S, Gyapong S, Farrokh A, Soergel P, Hillemanns P, Degenhardt F. Diagnostic performance and inter-observer concordance in lesion detection with the automated breast volume scanner (ABVS). *BMC Med Imaging*. 2013;13:36. PMID: 24219312. **KQ1E3c**

102. Yaghjian L, Colditz GA, Drake B. Vitamin D and mammographic breast density: A systematic review. *Cancer Causes Control*. 2012;23(1):1-13. PMID: 21984232. **KQ3E5**

103. Zuley ML, Guo B, Catullo VJ, Chough DM, Kelly AE, Lu AH, et al. Comparison of Two-dimensional synthesized Mammograms versus Original Digital Mammograms alone and in combination with Tomosynthesis images. *Radiology*. 2014;271(3):664-71. PMID: 24475859. **KQ1E3c**