Washington State Health Care Authority

Health Technology Assessment

Appropriate Imaging for Breast Cancer Screening in Special Populations

Final Evidence Report

December 10, 2014

Health Technology Assessment Program (HTA)
Washington State Health Care Authority
PO Box 42712
Olympia, WA 98504-2712
(360) 725-5126
hca.wa.gov/hta
shtap@hca.wa.gov

FINAL APPRAISAL DOCUMENT

Appropriate Imaging for Breast Cancer Screening in Special Populations

December 10, 2014

Daniel A. Ollendorf, PhD	Chief Review Officer
Anne M. Loos, MA	Research Associate
Jeffrey A. Tice, MD	Associate Professor of Medicine, University of California San Francisco
Janie M. Lee, MD, MS	Associate Professor of Radiology, University of Washington
Steven D. Pearson, MD, MSc	President

Table of Contents

About ICER

The Institute for Clinical and Economic Review (ICER) is an independent non-profit health care research organization dedicated to improving the interpretation and application of evidence in the health care system.

There are several features of ICER's focus and methodology that distinguish it from other health care research organizations:

- Commitment to aiding patients, clinicians, and insurers in the application and use of comparative effectiveness information through various implementation avenues, including its flagship initiatives, the New England Comparative Effectiveness Public Advisory Council (CEPAC; cepac.icer-review.org) and the California Technology Assessment Forum (CTAF; www.ctaf.org).

- Focus on implementation and evaluation of ICER research to create innovative decision support tools, insurance benefit designs, and clinical/payment policy.

- Deep engagement throughout the process with all stakeholders including patients, clinicians, manufacturers, purchasers, and payers.

- Inclusion of economic modeling in our research, and use of an integrated rating system for comparative clinical effectiveness and comparative value to guide health care decisions.

- ICER's independent mission is funded through a diverse combination of sources; funding is not accepted from manufacturers or private insurers to perform reviews of specific technologies. A full list of funders, as well more information on ICER's mission and policies, can be found at www.icer-review.org.

Acknowledgements

ICER would like to thank the following individuals for their expert opinion as well as peer review of draft documents:

Christoph Lee, MD, MSHS
Assistant Professor of Radiology, Breast Imaging Section
University of Washington School of Medicine
Assistant Professor of Health Services
University of Washington School of Public Health
Radiologist, Seattle Cancer Care Alliance

Jafi A. Lipson, MD
Assistant Professor of Radiology, Breast Imaging Section
Stanford University School of Medicine
Radiologist, Stanford University Medical Center

Executive Summary

Introduction

Breast cancer is the most common form of cancer in women (Siegel, 2013). An American woman is estimated to have a one in eight chance of developing invasive breast cancer at some time during her life. In 2013, there will be an estimated 234,580 new cases of breast cancer in the United States and an estimated 39,620 deaths from this cancer. This represents approximately 29% of all new cancer cases and 14% of all cancer deaths in women (Siegel, 2013). Moreover, breast cancer is the single leading cause of death for non-smoking women between the ages of 35 and 54 years, accounting for about 10% of all deaths (Woloshin, 2008).

Mortality from breast cancer has declined by about 2.2% per year since 1990, a 28% overall decline (Ries, 2008). The median values from a series of models estimated that a little more than half of the decline was due to improvements in therapy for breast cancer and that a little less than half (46%) was due to early diagnosis from mammography (Berry, 2005). This remains the dominant view, but a recent analysis of 30 years of data from the United States Surveillance, Epidemiology, and End Results (SEER) data called those conclusions into question (Bleyer, 2012). Bleyer and Welch estimated that 31% of breast cancer diagnosed with mammography represents "overdiagnosis" (i.e. , identification of cancers unlikely to cause significant morbidity or mortality) and concluded that screening mammography has had, at best, only a small effect on breast cancer mortality (Bleyer, 2012).

Screening for Breast Cancer

The primary method used to screen for breast cancer is mammography. Nine large clinical trials established the efficacy of screening mammography by randomizing over 600,000 women and following them for ten to twenty years (Alexander, 1997; Andersson, 1988; Andersson, 1997; Bjurstam, 1997a; Bjurstam, 1997b; Frisell, 1997a; Frisell, 1997b; Miller, 1992a; Miller, 1992b; Miller, 1997; Miller, 2000; Miller, 2002; Nystrom, 2002; Shapiro, 1997; Shapiro, 1988; Taber, 1995; Tabar, 1989; Tabar, 2000; Moss, 2006). The results have been summarized in many systematic reviews and meta-analyses (Swedish Cancer Society, 1996; Armstrong, 2007; Cox, 1997; Elmwood, 1993; Glasziou, 1995; Gotzsche, 2000; Hendrick, 1997; Humphrey, 2002; Kerlikowske, 1997; Kerlikowske, 1995; Nystrom, 1993; Rangash, 2001; Smart, 1995). There is general consensus that, for women between the ages of 50 and 69 years, screening mammography reduces breast cancer mortality by approximately 20% to 25% after 15 years of follow-up (Kerlikowske, 1997). For average-risk women between the ages of 40 to 49 years, there remains significant controversy about whether the benefits of routine mammography outweigh the harms, but most guidelines recommend either routine mammography or a discussion of the benefits and risks of mammography (USPSTF, 2009; Bevers, 2009; Lee, 2010; Smith, 2013).

Digital Mammography

Mammography was traditionally performed with film. It was one of the last radiographic procedures to transition from film to digital imaging because mammography requires extremely high resolution to be effective. Digital image acquisition improves the signal to noise ratio of x-ray detection over a wider contrast range than film (Feig, 1998; Pisano, 2000; Pisano, 1998). Digital enhancement of the images at computer workstations may also improve the accuracy of mammographic interpretation (Pisano, 2007). In particular, increased contrast resolution improves the detection of low contrast lesions in

radiographically dense breasts. Digital mammography has become the standard across the United States. As of July 1, 2013, 91.4% (11,705 / 12,800) of all U.S. mammography machines accredited by the Food and Drug Administration (FDA) are full-field digital (FDA, 2013).

Digital Breast Tomosynthesis (DBT)

Digital breast tomosynthesis (DBT) uses a conventional x-ray source that sweeps along an arc around the breast to acquire multiple two-dimensional (2-D) digital images (Houssami, 2013, Kopans, 2013; Sechopoulos, 2013). Breast compression is performed using the same device and technique as conventional mammography. The procedure to obtain each digital view is completed in less than 20 seconds. One of the advantages of DBT is that the images can be acquired immediately following the digital mammogram without needing additional compression. Like MRI, computational algorithms synthesize the resulting 2-D digital images to create tomograms (i.e., slices) allowing for a 3-D reconstruction of the breast. The tomograms can be displayed individually (similar to enhanced conventional mammograms) or in a dynamic movie mode.

The dose of ionizing radiation for DBT is about the same as that used for a conventional mammogram. A standard digital image was acquired in early DBT protocols, which served to increase the total radiation dose to approximately twice that of digital mammography alone (Housammi, 2013; Skaane, 2013). However, software has now been developed to produce a virtual 2-D mammographic image as part of the reconstruction of the tomosynthesis image, which may eliminate the need for a digital mammogram and thereby result in no increased radiation exposure. Early reader studies have shown comparable or better performance with so-called “two-view” DBT in comparison to digital mammography with one-view DBT or digital mammography alone (Zuley, 2014; Rafferty, 2014).

There are other uncertainties with DBT that should be considered in any discussion of its potential to replace digital mammography, however. First, the technology and algorithms used for DBT are still in evolution and have not yet been fully validated (Houssami, 2013, Kopans, 2013; Sechopoulos, 2013). One of the crucial areas is the development of techniques to biopsy lesions that are only seen on DBT (Vialai, 2013), although the FDA recently approved a device (Affirm®, Hologic, Inc.) that can be used with both mammography and DBT systems (FDA, 2013). In addition, the current reading time for DBT is about twice that required for digital mammography (Housammi, 2013; Skaane, 2013). Nevertheless, clinical interest in DBT is very strong. A recent publication summarized a 2012 survey of U.S. breast imaging centers (Hardesty, 2014); in the Western region of the U.S., one-third of centers were offering DBT, a percentage that has almost certainly increased in the two years since the survey.

Supplemental Screening Modalities for Breast Cancer Screening

There are many imaging approaches to screen for breast cancer in addition to mammography. Magnetic resonance imaging (MRI) has been most widely used. The American Cancer Society first recommended the use of MRI to screen women at highest risk for breast cancer in 2007, based primarily on genetic susceptibility (Saslow, 2007). Hand-held ultrasound has been used as a diagnostic tool to evaluate women with breast masses and has been promoted by some as a screening tool (Mahoney, 2013). The FDA recently has approved automated whole breast ultrasound, which scans and records ultrasound images of the entire breast, for breast cancer screening (Kelly, 2011; Giuliano, 2013). All of the advanced imaging technologies considered in this assessment generate multiple two-dimensional images representing slices of the breast. This allows the radiologist to visualize the breast in three-dimensions. This is particularly relevant in mammographically dense breasts because breast cancers may be obscured by superimposed dense tissue.

Magnetic Resonance Imaging (MRI) of the Breast

Magnetic resonance imaging uses strong magnetic fields to image the breast, rather than ionizing radiation. The system uses computational algorithms to generate detailed cross-sectional views of the breast. Mammography requires repositioning of the breast and mammography system for each desired view. In contrast, the MRI examination is typically performed with the patient in the prone position lying on a platform placed in the MR chamber that allows the breast to extend dependently from the patient and does not require repositioning. A contrast agent, gadolinium, is injected through an intravenous catheter (IV) to improve the images of the breast.

In studies of high-risk women, MRI approximately doubles the number of breast cancers that are detected compared to film mammography or breast ultrasound (Berg, 2004; Kagen, 2007; Kriege, 2004; Kuhl, 2005; Leach, 2005; Saradnelli, 2007; Tilanus, 2000; Warner, 2004). However, several factors limit the widespread use of MRI for screening. These include an increase in false positive test results, the need for placement of an intravenous catheter to infuse contrast, the length of time required for the examination, the cost of the examination, limited availability of breast MRI facilities (with special breast-specific magnetic coils and biopsy capability), and contraindications to the use of MRI due to pacemakers and other metallic implants. In addition, mammography has been found to be more sensitive than MRI for the detection of ductal carcinoma in situ (DCIS), a noninvasive cell abnormality in the milk ducts and some invasive breast cancers, so the two are typically used together (Bevers, 2009; Saslow, 2007; Mann, 2008).

Hand-held Ultrasonography (HHUS) of the Breast

HHUS is widely used at breast imaging centers to evaluate breast masses and to guide both cyst aspiration and percutaneous breast biopsy procedures. It is particularly useful to differentiate fluid filled cysts from solid masses (cysts are rarely cancerous). Over time, HHUS has evolved to use higher frequency sound waves to generate images of the breast with improved resolution. In addition, earlier generations of HHUS were not able to penetrate deeply into breast tissue and had a limited field of view. Advantages of ultrasound include the ability to evaluate tissue that is dense on mammography without additional ionizing radiation, which can potentially increase the risk for future cancers. It is also perceived to be more comfortable than mammography because it does not require compression of the breasts.

Ultrasound also has limitations. The primary concern with HHUS is the high number of false positive findings, which often lead to unnecessary biopsies. There are also concerns about the operator dependency and reproducibility of the examinations. Like MRI, HHUS takes time. The average length of time for breast HHUS imaging in a recent study was 19 minutes (Berg, 2008). In that study and many others, a breast radiologist performed the study. At a minimum, the breast radiologist needs to be available to review static images saved by the performing technologist in real time so that additional images can be acquired if necessary.

Automated Whole Breast Ultrasonography (ABUS)

ABUS uses computer driven ultrasound transducers to scan the entire breast under the guidance of a technician. A technician compresses the woman's breasts to her chest wall and applies ultrasound gel. A breast-shaped transducer is placed on the compressed breast and automatically scans the entire breast. The entire procedure, including patient preparation, takes about 15 minutes to complete (Kelly, 2011). ABUS reduces the need for radiologists to perform the scan and decreases the length of time of the exam, thus addressing two of the shortcomings of HHUS. It also produces a scan that should have less operator dependence. The radiologist can review the scan independently using software that displays the images individually or sequentially in a movie mode. The primary drawbacks to ABUS are

the inability to image very large breasts, the storage requirements for the data acquired during the scan, and the time required to read the scans (Kelly, 2011).

There are two major policy considerations surrounding the use of advanced imaging approaches in breast cancer screening. The first is the potential for DBT to replace digital mammography as a frontline screening tool in asymptomatic women. Because this is a new technology, the evidence base is expected to be limited, particularly with respect to long-term patient outcomes.

The other major consideration relates to the use of supplemental screening among women with a normal mammogram (i.e., no abnormalities detected) but with dense breast tissue that might obscure an abnormality. Breast density is qualitatively assessed by the radiologist based on mammographic images into one of four possible letter designations: (a) almost entirely fatty, (b) scattered areas of fibroglandular density, (c) heterogeneously dense, which may obscure small masses; or (d) extremely dense, which lowers the sensitivity of mammography (BI-RADS, 2013). The term "dense breast tissue" has primarily been applied to categories (c) and (d).

Supplemental screening is a generally-accepted practice among women with very strong risk factors for breast cancer, such as BRCA mutations or significant family history of the disease. However, these represent a small proportion of screened women. In contrast, dense breast tissue is present in nearly 50% of screening-age women (BI-RADS, 2013). While the presence of dense breast tissue has also been acknowledged as an independent (although modest) risk factor for breast cancer and denser tissue may mask tumors on standard mammography, little is known about the potential impact of supplemental screening if it were to be expanded to all women with dense breast tissue regardless of overall breast cancer risk.

Nevertheless, within the last 5 years, 19 states have passed legislation requiring physicians to notify women if they have dense breast tissue, largely as a result of patient advocacy efforts fueled by situations of missed cancer on mammography (Are You Dense Advocacy, 2014). Some of these mandates also require insurers to cover supplemental screening in such women. Many patient advocacy groups have commended these efforts, stating historically poor communication between the medical community and patients about the limitations of mammography (Lee, 2013). Others are concerned that such mandates are premature, as the current literature does not provide evidence of the benefits of supplemental screening in such a large and diverse population (D'Orsi, 2012). Advocates for DBT have also stated that the three-dimensional visualization may obviate the need for supplemental screening in women with dense breast tissue, but there are questions about whether there is sufficient evidence to support this claim. Payers and policymakers alike are concerned about the level of benefit that might be gained from supplemental screening in this population relative to the potential harms of patient anxiety, overdiagnosis and overtreatment, and false-positive findings.

Appraisal Scope

This appraisal focuses on the use of digital breast tomosynthesis (DBT) as a "frontline" screening test among all women who are candidates for breast cancer screening, in comparison to conventional digital or film mammography. In addition, evidence is assessed for other tests used as supplemental screening tests in women with dense breast tissue and a negative mammographic or DBT result, including handheld ultrasound, automated ultrasound, and magnetic resonance imaging. Key questions for the evaluation can be found below.

Key Questions

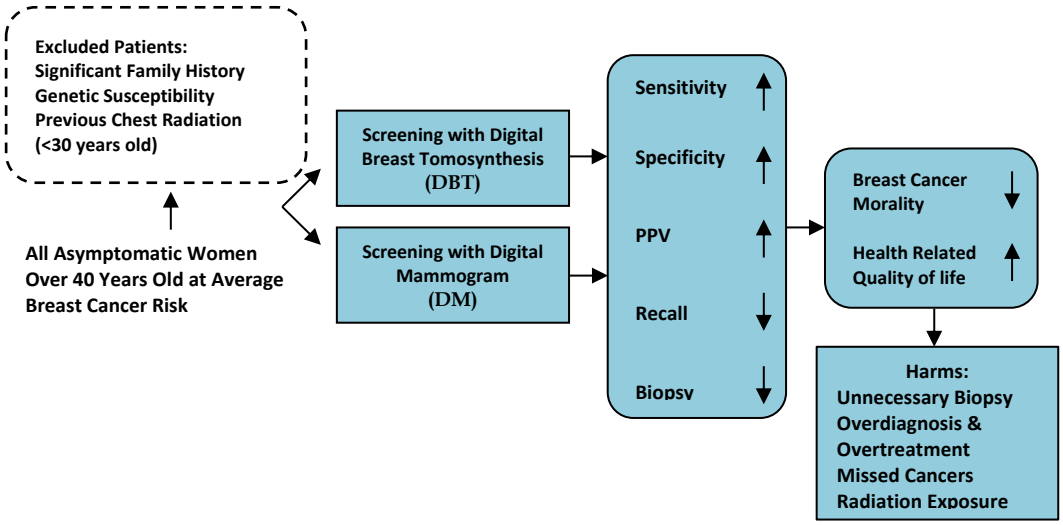
1. What is the effectiveness of screening with digital breast tomosynthesis (DBT) vs. digital mammography among women aged 40-74 who are candidates for screening mammography?
2. What is the comparative effectiveness of handheld ultrasonography, automated ultrasonography, and magnetic resonance imaging when used as supplemental screening modalities in women with dense breast tissue and a negative mammogram or negative DBT result?
3. What are the documented and potential harms associated with these imaging tests, including overdiagnosis and overtreatment, unnecessary biopsy as a result of false-positive imaging, patient anxiety, and radiation exposure?
4. What is the differential effectiveness and safety of the tests of interest according to such factors as age, race or ethnicity, comorbidities, BMI, method of breast density classification, overall breast cancer risk, scan vendor, and imaging protocol (e.g., whether ultrasound is performed by a radiologist, technologist, or some combination of the two)?
5. What are the costs and cost-effectiveness (e.g., cost per cancer detected) of the imaging modalities of interest?

Analytic Framework

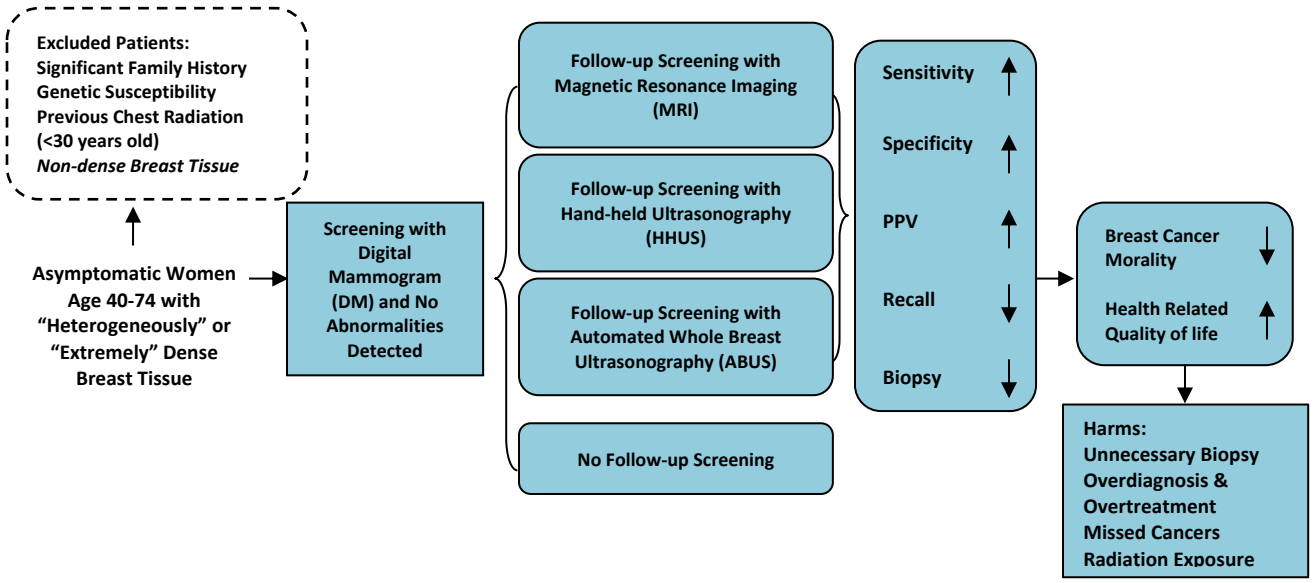
Two distinct analytic frameworks were used for this review, as shown in the figure on the following page. "Search A" relates to the use of DBT or digital mammography in in all women eligible for screening, and "Search B" depicts the conceptual flow in consideration of the evidence for the supplemental screening tests of interest among women with a negative mammogram or negative DBT result. Note that the figure is intended to convey the conceptual links involved in evaluating outcomes of the imaging modalities of interest, and is not intended to depict a clinical pathway through which all patients would flow. As is the case for many screening or diagnostic tests, it was expected that data linking screening modalities to direct patient outcomes will be limited, requiring instead a series of conceptual links between test characteristics and the major outcomes of interest.

Analytic Framework: Breast Cancer Screening

Search A



Search B



Study Quality

We used criteria published by the U.S. Preventive Services Task Force to assess the quality of RCTs and comparative cohort studies, using the categories “good”, “fair”, or “poor”. Guidance for quality rating using these criteria is presented below (AHRQ, 2008).

***Good**: Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention to treat analysis is used.*

***Fair**: Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done for RCTs.*

***Poor**: Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.*

In addition, the QUADAS-2 tool was used specifically to measure the quality of diagnostic accuracy studies (Whiting, 2011). Overall strength of evidence for each screening modality of interest was described as “high”, “moderate”, or “low”, and utilized the evidence domains employed in the AHRQ approach (AHRQ, 2012). The quality of the supplemental screening studies was evaluated using an adaptation of the QUADAS-2 specific to concerns regarding supplemental screening (see Table 6 on pages 44 through 50). Specifically, patient inclusion criteria were expanded to consider (a) the use of digital vs. film mammography as the initial test; and (b) the approach used to classify breast density (some studies accepted BI-RADS category 2 or “b” [scattered densities]). In keeping with standards set by the Washington HCA, however, assignment of strength of evidence will focus primarily on study quality, quantity of available studies, and consistency of findings.

Finally, summary ratings of the comparative clinical effectiveness and comparative value of the tests of interest (i.e., across multiple key questions) will be assigned using ICER’s integrated evidence rating matrix (Ollendorf, 2010). The matrix has been employed in previous Washington HCA assessments of virtual colonoscopy, coronary CT angiography, cervical spinal fusion for degenerative disc disease, cardiac nuclear imaging, and most recently, proton beam therapy.

Data from all retrieved studies were included in evidence tables regardless of study quality.

Results

Evidence Quality

The quality of available evidence on: (a) use of digital breast tomosynthesis versus digital mammography as a frontline general population screening tool; and (b) use of automated and handheld ultrasound as well as magnetic resonance imaging for supplemental screening in women with dense breast tissue, is summarized in the tables on pages 44 through 51. We identified a total of 33 studies. **Importantly, *none* of the studies identified assessed the impact of either general population or supplemental screening on breast cancer morbidity or mortality**. Nine studies evaluated the use of DBT in women undergoing routine screening. Of the remaining 24 studies examining supplemental screening in women with dense breast tissue and a negative mammogram, we identified 18, five, and one for HHUS, ABUS, and MRI respectively. Most of the comparative studies identified had major quality concerns. The majority of studies were observational in nature; no randomized control trials were available. Because the population of interest differed for each search, we used separate criteria to assess study quality.

All of the nine studies evaluating the use of DBT were rated "poor" based on the QUADAS-2 assessment, primarily because of no or incomplete follow-up; breast cancer screening studies should be at least as long as the interval between screening tests (i.e., 1-2 years) to adequately capture interval cancers. Only one study (Destounis, 2014) had adequate follow-up for evaluating interval cancers; however, patient groups were imbalanced with respect to breast cancer risk factors and other clinical characteristics in this study. While all studies included asymptomatic women presenting for routine screening, limited availability of DBT at screening sites and selection and/or volunteer bias (i.e., systematic differences between patients who agree to undergo a test and the target population for study) were issues in six of the studies (Destounis, 2014; Freidewald, 2014; Greenberg, 2014; Haas, 2013; McCarthy, 2014; Rose, 2013). The only two prospective studies (Ciatto, 2013; Skaane, 2013) had adequate reporting of nearly all outcomes of interest; number of biopsies as a result of abnormal findings was not included in either study, however. Study quality is summarized in Table 5 page 55.

The quality of the supplemental screening studies was evaluated using an adaptation of the QUADAS-2 specific to concerns regarding supplemental screening. The single study (Kuhl, 2010) evaluating the use of MRI in women with dense breasts was found to be of fair quality, as consecutive women were screened using the various MRI protocols, and two-year follow-up data were recorded in nearly all women (98%). Nevertheless, this was not a true screening study as nearly half of women had a personal history of breast cancer; in addition, applicability to U.S. settings is limited, as the majority of women had not only a negative mammogram but a negative ultrasound result as well.

Although we identified the most studies on the use of HHUS as a supplemental screening tool in women with dense breasts, only one study (Berg, 2012) was judged to be of good quality. The interval between the mammogram and the HHUS examination should be relatively short (i.e., with one month of each other). Otherwise the HHUS may find cancers that would also be visible on a mammogram at a later point in time. One study (Hooley 2012) included HHUS results from as much as 361 days after the mammogram – it is likely that mammography performed at that point would find additional cancers as well. Twelve studies did not report the time interval between examinations, one study reported that there was an average of two months between the examinations, and three studies performed both examinations within the same month.

Four of the five studies evaluating ABUS as a supplemental screening tool were of poor quality. There was a high degree of uncertainty for several outcomes, including biopsy and recall rate. The former was

not reported in most studies, and the latter was reported with significant variation across studies. It is likely that these inconsistencies are in part due to the high degree of technical proficiency required to perform the procedure, though only one study (Arleo, 2014) addressed this issue.

A summary evidence table capturing the strength of evidence for each of the five key questions of interest can be found in Table ES1 on the following page. As described at the beginning of this section, there were no studies identified that directly measured the impact of DBT or any of the supplemental screening modalities on breast cancer morbidity and mortality, so assessment of effectiveness is limited to diagnostic performance and cancer detection alone. A detailed assessment of the evidence for each study question is presented in the sections that follow.

Table ES1: Summary evidence table: Impact of DBT and supplemental screening with MRI, HHUS, and ABUS on key questions of interest.

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
KQ1: Effectiveness of General Population Screening with DBT vs. Digital Mammography								
N=313,298 RCT=0	DM	Medium	Consistent	Indirect	Imprecise	++ Low	Incremental to DM	Incomplete follow-up precludes definitive conclusions
KQ2: Effectiveness of Supplemental Screening with MRI, HHUS, and ABUS vs. Digital Mammography Alone								
MRI N=427*, 5,652 RCT=0	Mammography (DM or film)	Medium	Consistent	Indirect	Imprecise	++ Low	Incremental to mammography	Only one study in relevant population, others in high-risk women
HHUS N=96,002 RCT=0	Mammography (DM or film)	Low	Inconsistent	Indirect	Precise	++ Moderate	Comparable to mammography	Substantial between-study heterogeneity
ABUS N=28,093 RCT=0	DM	High	Inconsistent	Indirect	Imprecise	+ Insufficient		Small evidence base, substantial heterogeneity
KQ3: Harms of General Population and/or Supplemental Screening with Tests of Interest								
DBT, MRI, HHUS, ABUS	Mammography (DM or film)	High	Inconsistent	Indirect	Imprecise	+ Insufficient		General underreporting of harms
KQ4: Differential Effects of General Population and/or Supplemental Screening Tests in Key Subgroups								
DBT, MRI, HHUS, ABUS	Mammography (DM or film)	High	Inconsistent	Indirect	Imprecise	+ Insufficient		Extremely limited subgroup data across studies

Study Information	Comparators	Risk of Bias	Consistency	Directness	Precision	Strength of Evidence	Direction of Effect	Comments
KQ 5: Costs and Cost-Effectiveness of Screening Tests of Interest								
DBT	DM	Low	N/A	Indirect	Imprecise	+ Insufficient	$50,000-$100,000 per QALY gained	One model-based study, not in target pop'n
MRI					*No studies**			
HHUS N=15,000 RCT=0	Mammography (DM or film)	High	Consistent	Direct	Imprecise	++ Low	$60,000-$200,000 per add'l cancer detected	Limited to 3 accuracy-based studies, long-term costs not included
ABUS					*No studies*			

*Single study in women with dense breast tissue and negative mammogram; others in women at very high breast cancer risk

DM: digital mammography; DBT: digital breast tomosynthesis; MRI: magnetic resonance imaging; HHUS: handheld ultrasound; ABUS: automated whole breast ultrasound

Key Question #1: What is the effectiveness of screening with digital breast tomosynthesis (DBT) vs. digital mammography among women aged 40-74 who are candidates for screening mammography?

We identified a total of nine studies (Ciatto, 2013; Skaane, 2013a; Skaane, 2013b; Haas, 2013; Rose, 2013; Friedewald, 2014; Destounis, 2014; Lourenco, 2014; Greenberg, 2014; McCarthy, 2014) evaluating the use of DBT in comparison to digital mammography. The primary data are summarized in Table ES2 below. Two large prospective studies (Skaane, 2013; Ciatto, 2013) compared the test characteristics of digital mammography with and without DBT performed in the same patients on the same day. Three additional studies study (Destounis, 2013; Haas, 2013; Greenberg, 2014) compared two groups of patients, one screened with digital mammography alone and the other with digital mammography plus DBT. Finally, two retrospective, multicenter studies (Friedewald, 2014; Rose, 2013) and two single-center studies (Lourenco, 2014; McCarthy, 2014) used data from prior screening examinations before and after implementation of DBT. Table ES2 below summarizes the nine DBT studies that compare the use of DBT to digital mammography. These studies include asymptomatic women presenting for routine screening in various sites across the United States and Europe. Some findings of interest are not reported, as follow-up was of inadequate duration or is currently incomplete to capture information on interval cancers.

Table ES2: Studies comparing DBT to digital mammography for screening of asymptomatic women.

		DBT					Sensitivity		Specificity	
Study	Women, N	Recall rate/1,000	CDR /1,000	Bx rate /1,000	PPV1 %	PPV3 %	M %	DBT %	M %	DBT %
Ciatto 2013	7,292	42.9	8.1	NR	18.8	NR	66.1	100	95.5	96.6
Skaane 2013	12,621	61.1	8.0	NR	13.1	NR	62.6	82.1	93.8	94.6
Haas 2013	6,100	84.0	5.7	NR	6.8	NR	100	100	NR	NR
Friedewald 2014	173,663	91.0	5.4	19.3	6.4	29.2	NR	NR	NR	NR
Rose 2013	9,499	54.5	5.4	10.6	9.8	24.7	100	100	91.7	95.1
Destounis 2014	524	42.0	5.7	11.5	13.6	50	100	75	97.9	99.4
Lourenco 2014	25,498	93.4	5.4	17.4	6.2	23.8	NR	NR	91.1	94.0
Greenberg 2014	59,617	135.8	6.3	26.3	4.6	23.8	NR	NR	84.3	87.0
McCarthy 2014	18,220	103.7	5.5	17.7	6.2	25.4	NR	NR	NR	NR

M: Mammography; DBT: Digital Breast Tomosynthesis

CDR = cancer detection rate; Bx rate = biopsy rate PPV1 = positive predictive value of a positive test result; PPV3 = positive predictive value of biopsies actually performed

Skaane and colleagues (Skaane, 2013a; Skaane, 2013b) recently published initial results from a large series of patients evaluated with both digital mammography and DBT performed on the same day. The study evaluated DBT in 12,621 women presenting for routine screening mammography in Oslo, Norway

in 2011. DBT added an average of 10 seconds per view to the time required for mammography (40 seconds total). The reading time also increased, from 45 seconds for mammography to 91 seconds for mammography plus DBT. The total radiation dose increased from 1.58 mGy for digital mammography to 1.95 mGy for DBT (note: this was prior to the introduction of software allowing for creating of mammographic images during DBT scanning). According to standard practice in Norway, two radiologists independently interpreted the images for each woman and the potentially positive cases were reviewed at an arbitration meeting. Follow-up is not complete, but three interval cancers were identified during nine months of follow-up. These were not included in the statistics reported in the paper, but they have been counted as false negatives in the calculations performed for this assessment.

Addition of DBT decreased both false positives and false negatives. Thus DBT had higher sensitivity (82.1% compared to 62.6%) and specificity (94.6% compared to 93.8%) than digital mammography alone (both p<.001). The cancer detection rate increased from 6.1 to 8.0 cases per 1,000 examinations (p=.001) while the recall rate decreased from 67.2 to 61.1 per 1,000 (p<.001). The adjusted increase in cancer detection was 40% (Rate Radio [RR] 1.40, 95% CI 1.13 to 1.71, p<.001).

There are several issues that make it difficult to generalize the results of this study to the U.S. The standard of care in Norway is to have two radiologists interpret each mammogram and to have an arbitration meeting to review all positive results and decide which to call back. As noted earlier, this approach has a much lower recall rate than that observed in the United States (Hofvind, 2008). In addition, follow-up for interval cancers was incomplete.

A similar study (Ciatto, 2013) compared digital mammography to DBT in 7,292 women coming in for routine screening mammography at two population-based centers in Italy. As in Norway, two radiologists independently interpreted the images for each woman. However, if either was positive, the woman was recalled in this study while in Norway, there was a conference to decide who should be recalled. The Italians found that DBT had greater sensitivity (100% compared to 66.1%) and greater specificity (96.5% compared to 95.5%) compared to digital mammography (both p<.0001). This translated into a higher cancer detection rate (8.1 compared to 5.3 per 1,000 examinations, p<.0001) with a lower recall rate (42.9 compared to 49.5 per 1,000 examinations). As in the prior study (Skaane, 2013), there was no long-term follow-up, so the primary outcomes were the cancer detection rate and the recall rate. The investigators also did not report the biopsy rate in the study.

The primary concern with the Ciatto study is the lack of follow-up for interval cancers. This artificially raises the sensitivity of DBT to 100% and causes an overestimation of the specificity and negative predictive value as well. A recently published post-hoc analysis of the Ciatto study (Houssami, 2014) comparing outcomes for different screening strategies found six additional cancers after the first year of follow-up and estimated the interval cancer rate to be 0.82 cancers per 1,000 screens (95% CI: 0.30–1.79/1,000). However, because all participants received integrated mammography and DBT, it was not possible to determine the independent impact of DBT in reducing interval cancers, and the authors suggested that the interval cancer rate from this analysis be interpreted with caution.

In a recently published analysis (Destounis, 2014) of U.S.-based DBT experience, Destounis and colleagues compared results among patients choosing to undergo DBT plus digital mammography (n=524) to those from a sample set of randomly selected women (n=524) screened with digital mammography alone during the same timeframe (June - December 2011) at a facility in New York. The combination of DBT and digital mammography had a significantly lower recall rate (42.0 compared to 114.5 per 1,000 examinations, p<.0001) compared to digital mammography alone. DBT was also associated with a lower biopsy rate (11.5 compared to 22.9 per 1,000 examinations), a higher cancer

detection rate (5.7 compared to 3.8 per 1,000 examinations), and a higher positive predictive value for those undergoing biopsy (50.0% [3/6] compared to 16.7% [2/12]); these differences did not appear to be statistically tested, however. The population examined appeared to be very low risk, as only six cancers were detected among 1,048 women screened (0.6%). After one year of follow-up, two women in the digital mammography group had a cancer diagnosis, neither of which were interval cancers. There were four cancers detected in the DBT group, of which one was an interval cancer. However, nearly 20% of women in the study did not complete one year follow-up, so these results are incomplete at best. In addition, selection bias could not be ruled out, as no adjustments were made for differences in breast cancer risk between groups, and the DBT group had higher proportions of women with dense breast tissue, personal or family history of breast cancer, and atypia on prior biopsy.

Another retrospective comparative cohort study conducted at four sites in the U.S. was released online by Haas and colleagues on July 30, 2013 (Haas, 2013). They compared the recall rate and cancer detection rate at sites using digital mammography (n=7,058) to those at sites using DBT (n=6,100). All women presenting for screening mammography were included except those with breast implants or those with large breasts requiring tiled images. As in the prior studies, DBT decreased the recall rate (84 compared to 128 per 1,000 examinations, p<.01) but the cancer detection rate was not significantly increased (5.7 compared to 5.2 per 1,000 examinations, p=.70).

This retrospective study (Haas, 2013) also has several major methodological concerns. As with the Destounis study, the mammography and DBT groups were not well matched. Women in the DBT group were younger (55.8 years compared to 57.5 years), had more extremely dense breasts (5.6% versus 3.0%) and less fatty breasts (8.8% versus 13.8%), were more likely to have a personal history of breast cancer (5.5% versus 2.8%), and were more likely to have a first-degree relative with breast cancer (18.8% versus 15.9%). The investigators did not adjust for these differences in their primary analyses, but did present the results of logistic regression analyses adjusted for age, breast density, family history and personal history of breast cancer. In those analyses DBT was associated with a 35% reduction in the odds of recall (p<.0001). The investigators did not report biopsy rates, so it is not possible to determine whether the reduction in the recall rate translated into a similar reduction in breast biopsies. Finally, there was no follow-up for interval cancers so the sensitivity, negative predictive value, and specificity cannot be calculated.

The largest DBT study conducted to date was a retrospective multicenter study (Friedewald, 2014) that evaluated the screening performance of 13 U.S. academic and nonacademic breast cancer screening centers over two periods – one year prior to the introduction of DBT, and one year following. After adjusting for site differences, the investigators retrospectively evaluated a total of 454,850 examinations, of which 281,187 were screened with digital mammography alone, and 173,663 with a combination of digital mammography plus DBT. While the number of recalls was lower for DBT compared to digital mammography (91 versus 107 per 1,000 screens, p<.001), the biopsy rate was higher for DBT (19.3 versus 18.1 per 1,000 screens, p=.004). Twelve of the 13 sites increased cancer detection with DBT with an overall rate of 5.4 compared to 4.2 per 1,000 examinations for mammography alone (p<.001). The addition of DBT was also associated with a significantly higher PPV for recalls (6.4% compared to 4.3%, p<.001) and for biopsy (29.2% compared to 24.2%, p<.001).

Despite its size and impressive findings, this study nevertheless carries some significant limitations beyond those of its pre-post design. As with the prospective studies described above, the Friedewald study lacks long-term follow-up for interval cancers, and therefore cannot provide a full picture of test performance. In addition, clinical information was limited to data required for reporting to regulatory authorities, so there may have been heterogeneity in the two populations analyzed that could not be

accounted for. Finally, statistical calculations were based on screen-level rather than patient-level data so it was not possible to consider each record to be an independent observation, which limited the statistical adjustments that could be made.

A newly-published retrospective multicenter study (Greenberg, 2014) compared 38,674 patients who underwent digital mammography to 20,943 patients who voluntarily selected DBT over a 16-month period. Although the study site description is not definitive, it appears that at least one of the sites also participated in the Friedewald study, so there may be some data overlap.

This study found a significant reduction in the recall rate with DBT (135.8 vs. 161.5 per 1,000 examinations), which represents a 2.6% absolute reduction in recalls (p<.0001). The cancer detection rate also increased from 4.9 in the DM group to 6.3 per 1,000 examinations in the DBT cohort, representing a 29% overall increase (p=.035). As with the Friedewald analysis, this study also reported a higher biopsy rate for DBT (26.3 vs. 21.6 per 1,000 examinations, p=.0003), and a PPV3 value that was similar for both groups. Although there were no differences in baseline characteristics between the two groups with regard to age, family history of breast cancer, or breast density, the investigators did find that significantly fewer DBT patients had less than two mammographic views at recall than those in the DM group (74.1% vs. 51%, p<.001), which may suggest that DBT is more effective at identifying patients who require additional follow-up.

Although this was one of two retrospective DBT studies to evaluate contemporaneous screenings, there are some limitations that may impact its findings. First, not only was there the potential for volunteer bias in the selection of DBT, most women were required to pay a $50 premium for the procedure, which may have introduced additional bias in the selection of candidates for DBT. In addition, this study assessed patients immediately following implementation of DBT at various times in several sites across three states; there were no adjustments to account for a learning curve with the new technology or potential heterogeneity within the study population.

A new publication of a retrospective study that also involved one of the participating sites in the Friedewald study (McCarthy, 2014) used a pre-post design to assess 26,299 screenings in 18,220 asymptomatic women over a 17-month period, and found a reduction in the recall rate with DBT that was statistically-significant in both unadjusted and adjusted models (8.8% vs. 10.4% for DM, OR 0.80, 95% CI 0.74-0.88, p<.001). Although this study was not powered to detect significant differences in biopsy rates (1.8% vs. 2.0% for DBT) or cancer detection rates (4.6 vs. 5.5 per 1,000 examinations for DBT) in the overall population, for women under age 50 there was a significant increase in the number of cancers detected with DBT compared to digital mammography (5.7 vs. 2.2 per 1,000 examinations, p=.02).

Another recent retrospective comparative cohort study (Lourenco, 2014) assessed 25,498 patients over one year before and after the implementation of DBT at a breast imaging center. In line with some other recent studies of DBT, biopsies were recommended more frequently in the DBT group (17.4 vs. 16.3 per 1,000 examinations for DM). There were also 31% fewer recalls with DBT (6.4% vs. 9.3%, p<.00001). It is unclear why absolute rates of recall were lower in this study than in others, given that it was U.S.-based. In contrast to recall findings, neither PPV3 nor the cancer detection rate differed significantly between groups. Finally, as with most of the other studies in this evaluation, follow-up was insufficient to detect interval cancers.

The final study of DBT (Rose, 2013) also used a pre-post design rather than a direct comparison of the two technologies. The investigators compared the screening benchmarks for the combination of DBT

and digital mammography (n=9,499) to those of digital mammography alone at the same sites in Texas during the prior year (n=13,856). There was no follow-up for interval cancers, so the sensitivity, specificity and negative predictive value are overestimated. As in the prior studies, DBT had a lower recall rate (54.5 compared to 87.2 per 1,000 examinations, p<.001). The biopsy rate was also lower (10.6 compared to 15.2 per 1,000 examinations) and cancer detection rate was higher (5.4 compared to 4.0 per 1,000 examinations), but these differences were not statistically significant.

Summary: Screening DBT

A total of nine studies of women presenting for routine screening for breast cancer found that DBT substantially decreased the recall rate relative to mammography, and most also found an increase in cancer detection. Evidence was more mixed on the biopsy rate, however, with newer studies showing an increased biopsy rate with DBT. Findings were also consistent across both the dense and non-dense subgroups in the studies reporting on those subgroups.

However, there were issues of study heterogeneity as well as comparability of screening populations. As shown in Table ES3 below, there is substantial uncertainty with recall rates, since the two prospective studies come from outside the U.S. where patterns of recall differ markedly. In addition, test performance is likely to be overstated in all of these studies, as follow-up is not long enough in any one study other than the Destounis study, and 20% of women in that study did not achieve one year of follow-up (Destounis, 2014).

Table ES3: Estimated yield of DBT in combination with digital mammography vs. digital mammography alone in women presenting for general population screening.

Statistic	Digital mammography	DBT+Digital mammography	Uncertainty
Recall rate per 1,000	100-160	80-140	Moderate-high
Biopsy rate per 1,000	14-22	12-27	Moderate
CDR per 1,000	3-5	4-6	Moderate-high
PPV3	20-25%	25-30%	Low-moderate

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

Key Question #2: What is the comparative effectiveness of handheld ultrasonography, automated ultrasonography, and magnetic resonance imaging when used as supplemental screening modalities in women with dense breast tissue and a negative mammogram or negative DBT result?

Screening Magnetic Resonance Imaging (MRI) of the Breast

The search identified only one study that evaluated the benefit of MRI following negative mammography in women with dense breasts. A recently published prospective study (Kuhl, 2014) evaluated the use of an abbreviated protocol (AP) form of MRI, consisting of a pre- and post-contrast acquisition (called MIP and FAST images, respectively) in breast cancer screening to evaluate its ability to detect cancers in asymptomatic women presenting for screening in Germany; findings were compared with a traditional full diagnostic MRI protocol (FDP). Women with heterogeneously dense or extremely dense breasts and a normal or benign mammography result as well as a negative or benign ultrasound (n=427) were included. All women were categorized according to their individual risk of developing breast cancer (i.e., mild, moderate, or high).

Acquisition time for AP was 184 seconds versus 1,024 seconds for FDP; reading time for an MIP image was 2.8 seconds and less than 30 seconds for AP. Both AP and FDP identified 11 cancers representing an additional cancer yield of 18.5 per 1,000 screens with 10 of the 11 cancers being identified using MIP images alone. Specificity (94.3% versus 93.9%) and positive predictive value (24.4% versus 23.4%) were not statistically different between AP and FDP readings, and negative predictive value was 100% for both protocols (95% CI, 99.3-100). No additional cancers were detected in the second screening round.

Although well-matched with regard to study population, the Kuhl study has several limitations. First, MRI was used as a third-line screening in women following a normal mammogram and negative or benign ultrasound result, which does not reflect current practice in the U.S. Nearly half of screened women had a personal history of breast cancer (220/443, 49.6%), so the high cancer detection rate likely reflects this underlying risk in the women referred for supplemental screening. Finally, outcomes from this study may not be applicable to screening in community clinical practice, as the radiologists assigned to review the images were considered expert readers.

Magnetic Resonance Imaging (MRI) for Screening High-risk Women

The majority of studies we identified from our search evaluated the use of MRI in women at a high risk of developing breast cancer, without regards to breast density. In particular, several large studies have evaluated the test characteristics of MRI in conjunction with mammography and ultrasound in BRCA1 and BRCA2 mutation carriers and other women at very high risk for breast cancer. Those studies are summarized briefly below as an update to the HCA review of MRI in such women conducted in 2010 (Delfini Group, 2010). The ACRIN 6666 study offered MRI to high-risk women who completed the third round of annual screening ultrasound and mammography in that study. The results will be described in the section on HHUS below.

Magnetic resonance imaging (MRI) has been primarily studied for breast cancer screening in women deemed to be at high risk either by personal history, family history or because they were known carriers of either a BRCA1 or BRCA2 mutation (Berg, 2004; Hagen, 2007; Kriege, 2004; Kuhl, 2005; Leach, 2005; Sardanelli, 2007; Tilanus-Linthrost, 2000; Warner, 2004; Kriege, 2006; Kuhl, 2000; Lehman, 2005; Lehman, 2007; Podo, 2002; Morris, 2003; Port, 2007; Stoutjesdijk, 2001; Yu, 2008; Warner, 2001; Trecate, 2006). These women have a lifetime risk greater than 20%, rather than the 10% to 20% lifetime

risk for most women with high breast density. No studies have demonstrated that MRI reduces the risk of death from breast cancer; there are no studies comparing women screened with MRI to other women screened with mammography alone and none of the studies of the test characteristics of MRI are of sufficient duration or size to evaluate patient-oriented outcomes such as the breast cancer recurrence or death from breast cancer.

Table ES4 on page 20 summarizes the larger prospective screening studies (n=5,652) that compare the use of MRI in high-risk woman to mammography with or without ultrasound. Women in these studies were primarily BRCA1 or BRCA2 mutation carriers or their first degree relatives. None of the studies followed women for more than one or two years. In addition, the majority of these studies compared MRI to film mammography only, since digital mammography was not widely disseminated until after publication of the DMIST trial in 2005.

The sensitivity of MRI for breast cancer in Table ES4 ranged from 77% to 100%. The sensitivity of mammography (25%-59%) and ultrasound (13%-65%) in these studies was about half that of MRI. In the largest three studies (Kriege, 2004; Kuhl, 2005; Leach, 2005), which included 52% of the cancers in all 14 studies, the sensitivity of MRI ranged from 71% to 91% while the sensitivity of mammography ranged from 32% to 40%. However, the specificity of MRI is consistently lower than mammography. In the same three studies, the specificity of MRI ranged from 81% to 97% compared to 93% to 99% for mammography, and in each individual study the specificity of MRI was lower than that of mammography. Because breast cancer is relatively uncommon, even in these high-risk populations, the lower specificity of MRI translates into a much higher number of false positive results. One study (Sardanelli, 2007) suggested that the high false positive rate decreases after the initial MRI. In that study the rate of false positive results declined from 14% initially to 8.2% on subsequent MRI's, but was still substantially higher than the 4.6% false positive rate for mammography (Kriege, 2006).

The cancer detection rate of MRI ranged from eight to 36 per 1,000 examinations in these studies – much higher than the three to six per 1,000 examinations typically reported in studies of mammography. This reflects in part the higher sensitivity of MRI and in part the higher incidence of breast cancer in these high-risk women.

However, this higher cancer detection rate comes at a cost: the biopsy rates in the MRI studies in Table ES4 range from 29 to 157 biopsies per 1,000 examinations. The biopsy rates are lower in studies of screening mammography (10 to 25 per 1,000 examinations). The PPV3 ranged from 17% to 89%, but the median was 48%, which is a very high yield per biopsy.

It is worth noting that the sensitivity of mammography and ultrasound were similar to each other in each of the five studies that report the sensitivity of all three screening technologies. The sensitivity of mammography and ultrasound in these studies is much lower than the sensitivity usually reported for these tests. The low sensitivity is due to the large number of cancers that are found by MRI alone – more than typically appear as interval cancers in the year following a screening examination. This suggests that many of the cancers detected by MRI would not have been diagnosed without MRI for more than one year after the examination. Early detection of cancers that would have become clinically apparent at a later date should translate into a higher cure rate and the need for less aggressive therapies, but some proportion of the cancers detected by MRI are likely to represent overdiagnosis – cancers that never would have become symptomatic in a woman's life.

The two systematic reviews described in Section 5 (Previous Systematic Reviews and Technology Assessments) both found that the addition of MRI significantly increased the sensitivity of screening for

breast cancer, but increased false positive results; the effect on breast cancer mortality remained unknown because none of the studies had sufficient follow-up duration to evaluate this endpoint (Lord, 2007; Warner, 2008). In one of the meta-analyses (Warner, 2008), adding MRI to mammography increased the sensitivity from 39% to 94%, but decreased specificity from 94.7% to 77.2%. If the prevalence of breast cancer in a high-risk population is 4.4% (the pooled prevalence across the 14 studies), then adding MRI to mammography in 1,000 women would detect an additional 24 breast cancers (increased from 17 to 41) and an additional 167 women would receive false positive results (increased from 51 to 218).

Table ES4: Prospective studies comparing magnetic resonance imaging, ultrasound, and mammography to screen high-risk women for breast cancer.

Study	Women, N	CDR MRI /1,000	Bx rate /1,000	PPV1 MRI %	PPV2 MRI %	PPV3 MRI %	Sensitivity			Specificity		
							M %	HHUS %	MRI %	M %	HHUS %	MRI %
Tilanus-Linthorst 2000	109	28	46	-	60	60	0*	-	100			
Podo 2002	105	67	86	-	89	89	13	13	100			
Kriege 2004	1909	12	29	-	57	57	40	-	71	95	-	90
Warner 2004	236	30	157	-	46	46	36	33	82	99	96	81
Kuhl 2005	529	36	147	-	50	50	32	40	91	97	91	97
Leach 2005	649	29	-	-	-	25	40	-	77	93	-	81
Lehman 2005	367	8	63	-	17	17	25	-	100	98	-	93
Lehman 2007	171	23	82	-	43	43	33	17	100	91		79
Sardanelli 2007	278	22	90	-	60	60	59	65	94	99	98	98
Kuhl 2010	687	15	34	-	48	48	33	37	93	99	98	98
Berg 2012	612	15	70	-	-	19	31	-	88	92	-	76

M: Mammography; HHUS: Ultrasound; MRI: magnetic resonance imaging

CDR = cancer detection rate

Bx rate: biopsy rate

PPV1 = positive predictive value of a positive test result (BI-RADS assessment 0, 4, or 5)

PPV2 = positive predictive value of a biopsy recommended (BI-RADS assessment 4 or 5)

PPV3 = positive predictive value of biopsies actually performed

Summary: Screening MRI of the Breast

There is very limited data evaluating MRI in a general screening population with dense breasts, as well as in populations at intermediate risk (15% to 20% lifetime risk). We identified only one study evaluating MRI as a supplemental screening tool in women with dense breasts and negative initial mammography and ultrasound findings. The results from this study show that an abbreviated version of MRI may hold some promise in detecting the presence of additional cancers without sacrificing levels of sensitivity and specificity associated with the full diagnostic protocol, with the potential for cost savings from the abbreviated protocol. However, this study was conducted in Europe, with screening and referral patterns that are not generalizable to the U.S. setting.

The data from high-risk populations suggest that the addition of MRI would more than double the cancer detection rate (best estimate 2.4-fold increase) with a four-fold increase in the recall rate (best estimate 4.3-fold increase). Estimates based on these data are shown in Table ES5 below. There is a high level of uncertainty around these values because of the lack of direct evidence from studies of MRI in women with dense breast tissue and because of the heterogeneity of the findings in the studies of high risk women summarized in ES5 below.

Table ES5: Estimated incremental yield of MRI after negative digital mammography in women with dense breast tissue.

Statistic	Digital Mammography	Incremental Yield With MRI	Uncertainty
Recall rate per 1,000	100-120	100-120	High
Biopsy rate per 1,000	14-22	20-40	High
CDR per 1,000	3-5	3-11	High
PPV3	20-25%	22%-48%	High

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

These estimates suggest that MRI would find substantially more cancers than those found by digital mammography, with a PPV3 between 22% and 48%. There would be approximately 100-120 additional recalls and between 20 and 40 additional biopsies per 1,000 women in order to identify approximately 3-11 additional cancers.

Screening Hand-held Breast Ultrasound (HHUS)

Eighteen studies of almost 100,000 women screened with HHUS met the search criteria for this assessment and are described in Table A1 in Appendix B, with a quality assessment of these studies presented in Table 6 beginning on page 44 (Buchberger, 2000; Kaplan, 2001; Kolb, 2002; Crystal, 2003; Leconte, 2003; Corsetti, 2008; Kolb, 2007; Maestro, 1998; Corsetti, 2006; Brancato, 2007, De Felice, 2007; Corsetti, 2001; Hooley, 2012; Leong, 2012; Weigert, 2012; Chae, 2013; Girardi, 2013; Parris, 2013; Venturini, 2013; Korpraphong, 2014). In general, all participants in these studies underwent mammography first and those with negative mammograms were subsequently screened by HHUS. One study by Corsetti and colleagues is presented twice in the tables: their 2008 publication (Corsetti, 2008) had a large number of examinations; and their 2011 publication (Corsetti, 2011) included one year follow-up for a subset of the women. Results from the ACRIN 6666 trial (Berg, 2008; Berg, 2012) are also described in the tables, although the study did not meet the inclusion criteria. However, it was the

only prospective study in the United States with complete reporting of the data on the combination of mammography and HHUS with one-year follow-up after more than one round of screening.

As shown in Table A1 in Appendix B, the participants in these studies had a mean age usually in the 50s with a broad range (25 to 91 years). Most included asymptomatic women presenting for screening mammography who were found to have dense breasts, although the definition of high density varied somewhat. The majority of the trials were done outside of the United States. Three recent retrospective cohorts (Hooley 2012; Weigert 2012; Parris 2013) described the findings in Connecticut, which was the first state to pass a law requiring breast density notification. These three studies represent the best evidence in the US population for the incremental cancer detection rate with HHUS, although they do not include any data on the interval cancer rate. The two other trials in the U.S. (Kaplan 2001; Kolb 2002) reported results from imaging performed in the year 2000 and earlier. A radiologist performed the HHUS in the majority of the studies. Nine of the studies reported no follow-up on participants, three reported variable follow-up on a subset of patients, and three reported one-year follow-up. This is typical for publications of data from mammography facilities, as they keep records on the follow-up of abnormal tests and cancer detection for quality assurance work, but do not routinely follow patients with normal mammography results to identify interval cancers. This also means that the sensitivity, specificity, and negative predictive value reported from those studies will overestimate the true values. The diagnostic accuracy test characteristics from these studies are summarized in Appendix B. In these studies, when one participant was diagnosed with more than one cancer or had more than one biopsy, the statistics were reported on a per participant basis rather than per cancer or biopsy. The statistics in Table A2 in Appendix B represent only participants who had a negative mammogram assessment and fell into one of the two high density BI-RADS categories (heterogeneously dense or extremely dense) except for those with separate rows for mammography and mammography plus supplemental screening.

Only four of the studies in Appendix Table A2 (Hooley 2012; Leong 2012; Girardi 2012; Parris 2013) compared HHUS to digital mammography. Ten studies compared HHUS to film mammography and one did not report the type of mammography machine used in the study. The ACRIN 6666 Trial used a mix of digital and film mammography (Berg, 2008; Berg, 2012).

Only three of the trials (Kaplan 2001; De Felice 2007; Leong 2012) in women with dense breasts and the ACRIN 6666 trial (Berg, 2008; Berg, 2012) were prospective studies. Prospective studies are more likely to have complete and consistent measurement of the key outcomes because they are defined objectively at the start of the study and collected systematically. It is worth noting in Table A2 that these trials had by far the highest recall rates (>100 recalls per 1,000 examinations). Most of the other studies did not systematically report recalls after ultrasound and often reported the HHUS assessment as positive only if a biopsy was recommended, thus underestimating the true recall rate.

The recall rate for HHUS after normal mammography ranged from 21 to 170 per 1,000 examinations with the median value across the studies of 59, lower than the typical recall rate for mammography described earlier in this report of 100 per 1,000 examinations. In the ACRIN 6666 study (Berg, 2012), HHUS recalled 186 women per 1,000 examinations. As noted above, all of the prospective studies (Kaplan, 2001; De Felice, 2007; Leong, 2012; ACRIN 6666) reported recall rates greater than 100 per 1,000 examinations, so these values are likely to be more accurate.

The biopsy rate for women having HHUS after normal mammography ranged from 12 to 114 per 1,000 examinations with a median of 46. In the ACRIN 6666 study the biopsy rate was 88 per 1,000 examinations the first round and about 61 per 1,000 examinations the third round. The cancer

detection rate varied from 0.4 to 14.2 per 1,000 examinations with a median value of 3.2 per 1,000 examinations. In the ACRIN 6666 trial, HHUS detected 5.9 cancers per 1,000 examinations. It is not clear why the biopsy rates vary across such a wide range. Potential explanations include incomplete reporting of cyst aspirations, different thresholds for performing cyst aspirations, operator dependency in performing HHUS, and differences in the proportion of patients undergoing a first time screening HHUS compared to those with prior examinations for comparison.

There was also a wide range of estimates across the studies for the PPV1 (2.0 to 11.7%, median 6.5%) and the PPV3 (3.2 to 18.4%, median 7.1%). The heterogeneity of these results was likely due to a combination of factors. These include the study design (prospective, retrospective), the use of film or digital mammography, differences in the assessment of mammography across countries, whether a radiologist or a technician performed the HHUS, the level of experience and training of the person performing the HHUS, and differences in the populations studied (age distribution, breast cancer risk factors, time since last mammogram).

The characteristics of the cancers detected by mammography alone and of ultrasound among women with a negative mammogram are described in Table A3 in Appendix B. The table shows that most of the cancers detected by HHUS after negative mammography are small, node negative, early stage cancers. These are the cancers that are potentially curable by early detection before they develop into cancers with a poorer prognosis. Cancers at an early stage also require less aggressive therapy: the patient may be eligible for lumpectomy rather than mastectomy and may not require systemic chemotherapy. Thus early detection may improve both quality and quantity of life. The counter-argument is that some of these early stage cancers may not have progressed much before the next routine screening examination with mammography. Thus, they may ultimately have been detected and cured with mammographic screening alone. In addition, some proportion of these cancers may represent overdiagnosis: the identification of a cancer that would not have ever progressed to cause symptoms prior to the death of that individual woman. The identification of such cancers would lead to unnecessary labeling of the woman as someone who has cancer as well as unnecessary surgery and chemotherapy. The only way to test which of these two competing hypotheses is true would be to perform a randomized trial comparing the two approaches to breast cancer screening.

A large retrospective study of asymptomatic women with dense breasts (n=20,864) presenting for routine screening underwent mammography (n=12,505) or mammography plus HHUS (n=8,359) in Korea over a two-year period (Chae, 2013). Screening with both ultrasonography and mammography increased the cancer detection rate (2.9 compared to 0.5 per 1,000 examinations) but also increased the recall rate (55 per 1,000 compared to 42 per 1,000). Of note, while no details were provided on abnormalities were adjudicated, the low cancer detection rates reported here are reflective of the much lower incidence of breast cancer in Asian countries relative to Western nations.

Of the 24 cancers identified with HHUS, 23 were invasive and one was DCIS. Eleven cancers were identified in the mammography-only group: five were false-negatives which were all subsequently identified with diagnostic ultrasonography, five were DCIS, and one was an invasive cancer. As a result, sensitivity of ultrasound was 100.0% compared to 54.5% for mammography alone (p=.002). PPV1 values were also significantly higher for HHUS (5.3% versus 1.1% for mammography, p<.001). However, false-positives were higher in the HHUS group (5.18%) compared to mammography alone (4.14%) and the PPV2 value was also much lower for HHUS plus mammography (11.1% [24/216] versus 50%, [6/12]) than for mammography alone; neither of these differences were tested for statistical significance.

There are several limitations to this study. While a broad spectrum of breast cancer risk was allowed in this study, ultrasonography was chosen voluntarily, women who had an elevated risk may have disproportionately opted for additional screening, thereby which may have artificially increased the cancer detection rate for ultrasound. In addition, women who opted for HHUS were asked to pay out of pocket for the test, which may explain why only 40% of women with dense breasts chose HHUS. Finally, this study does not fully meet criteria because there was no requirement for a negative mammogram before the decision for supplemental screening with HHUS was made.

Another recent study (Korpraphong, 2014) evaluated of the use of hand-held ultrasound in asymptomatic women with non-fatty breasts (BI-RADS 2, 3, or 4 density) who presented for mammography screening in Thailand and were subsequently examined with HHUS. Subgroups of women were analyzed according to age and breast density. Of the 14,483 screenings, 115 cancers were detected: 31 with mammography alone, 19 with HHUS alone, and 55 with both HHUS and mammography. The overall incremental cancer detection rate was 1.4 per 1,000 examinations ($p<.0001$). Women with extremely dense breasts and women between 40-59 years old had an incremental cancer detection rate of 2.5 and 2.0 per 1,000 examinations, respectively, which were the highest among all categories; statistical significance was not reported for either of these findings, however. In addition, no outcomes were assessed for biopsy rate, recall rate, or interval cancers. The investigators also warn there is a high incidence of breast cancer in the community where the study was conducted, making it difficult to apply these results to the general population.

American College of Radiology Imaging Network (ACRIN) 6666 Trial

Because the ACRIN 6666 trial (Berg, 2008; Berg, 2012) was the only prospective trial performed in the United States with mostly digital mammography and one year follow-up for multiple screening rounds (good quality), its findings are the most pertinent to the focus of this review and will therefore be described in detail below. The population studied was higher risk than that of a typical screening population, so the biopsy rate, cancer detection rate, and positive predictive values will be higher than those of a screening population. For instance, in the first round the biopsy rate based on mammography was 14.4 per 1,000 examinations, the cancer detection rate was 7.5 per 1,000 examinations, and the PPV3 was 31%, all of which are higher than expected for mammography in a screening population (approximately 10 per 1,000, 5 per 1,000, and 25% respectively).

The ACRIN 6666 trial randomized 2,809 high-risk women to receive both mammography (film or digital) and ultrasound in alternate order (Berg, 2008). High-risk was defined by at least one of the following: a personal history of breast cancer; positive for BRCA1 or BRCA2 mutation; a lifetime risk ≥ 25%, a 5-year risk ≥ 2.5% or ≥ 1.7% with extremely dense breast tissue; prior biopsy with atypical ductal hyperplasia, atypical lobular hyperplasia, lobular carcinoma in situ or atypical papilloma; or prior mantle radiation. The study also required that the women have at least one quadrant of one breast with heterogeneously dense or extremely dense tissue on a prior mammogram. The trial did not meet the inclusion criteria for this assessment for two reasons: the study subjects are high risk rather than a general screening population and were not required to have dense breasts by BI-RADS criteria.

The women were followed for three annual cycles and upon completion of the third cycle, the women were offered additional screening with breast MRI (Berg, 2012). There were 2,659 women with data for analysis after the first year of follow-up. Their median age was 55 years and 93% were white. The primary risk factors for inclusion in the study were a personal history of breast cancer (53%), a lifetime risk ≥ 25% (19%), and a five-year risk ≥ 2.5% (15%). The investigators present the results of mammography alone and for the combination of mammography plus ultrasound, but not for ultrasound

alone or the subgroup of women with a negative mammography assessment. When possible, we calculated the incremental results for ultrasound following negative mammography.

In the first screening round, mammography detected 20 cancers (cancer detection rate 7.6 per 1,000 examinations) and ultrasound detected an additional 14 cancers (5.9 per 1,000 examinations) (Berg, 2012). There were two interval cancers so the sensitivity of mammography was 55.6% (20/36) and the sensitivity of ultrasound in women with negative mammograms was 87.5% (14/16). The number of recalls increased from 306 with mammography alone to 707 with mammography plus ultrasound, a 2.3-fold increase in the recall rate (from 115.1 per 1,000 examinations to 265.9 per 1,000). The number of breast biopsies increased from 65 to 272, a 4.2-fold increase (from 24.4 per, 1,000 examinations to 102.3 per 1,000). The PPV3 for ultrasound in women with negative mammograms was only 6.8%.

By the third screening examination, the test characteristics changed, reflecting a reduction in prevalent cancers due to early detection, the transition to digital mammography, and improved specificity with increased experience of the radiologists and the availability of prior examinations available for review (Berg, 2012). Mammography detected 23 cancers (cancer detection rate 9.9 per 1,000 examinations) and ultrasound detected an additional nine cancers (4.2 per 1,000 examinations). There were 14 interval cancers so the sensitivity of mammography was 50.0% (23/46) and the sensitivity of ultrasound in women with negative mammograms was 39.1% (9/23). The investigators did not report the recall rate and biopsy rate for round three, but did report the numbers for the combination of rounds two and three. The number of recalls increased from 453 with mammography alone to 809 with mammography plus ultrasound, a 1.8-fold increase in the recall rate (from 94.1 per 1,000 examinations to 168.1 per 1,000). The number of breast biopsies increased from 97 to 339, a 3.5-fold increase (from 20.1 per 1,000 examinations to 70.4 per 1,000). The PPV3 for ultrasound in women with negative mammograms was 7.1%.

In round three, women were offered MRI in addition to HHUS and mammography (Berg, 2012). The 612 women in the MRI sub-study had higher risk for breast cancer and were younger than those who declined participation (Berg, 2010). In this group of participants, mammography alone detected five cancers, ultrasound detected an additional two cancers (sensitivity for the combination 43.8%, cancer detection rate 11.4 per 1,000 examinations) and MRI detected nine additional cancers (sensitivity 100%, incremental cancer detection rate 14.7 per 1,000 examinations and combined cancer detection rate 26.1 per 1,000 examinations). The nine cancers detected by MRI only were small (median 8.5 mm) and all were lymph node negative. Both cancers seen only with HHUS (not mammography) were also diagnosed with MRI. The high cancer detection rate in the women in the MRI group reflects the high underlying risk for cancer in the women who agreed to participate in the sub-study. The recall rate was 85.0 per 1,000 examinations for mammography alone, 163.4 per 1,000 for the combination of mammography plus HHUS and 260.0 per 1,000 for MRI. The biopsy rate was 62.1 per 1,000 examinations for the combination of mammography plus HHUS and 132.3 per 1,000 for the combination with MRI. The PPV3 for MRI in women with a negative mammogram was 22.4%, which is much higher than that of ultrasound.

In this high-risk population, the ACRIN 6666 study found that supplemental screening with HHUS produced a relatively high yield of cancers the first round of screening, approximately doubling the cancer detection rate, but this decreased with subsequent rounds. In order to find these cancers, the recall rate more than doubled so that one in four women (26.6%) were recalled in the first round. The number of biopsies performed increased by a factor of four. In the first round, the combination of ultrasound plus mammography led to almost as many biopsies (10.2% of women) as women recalled with mammography alone (11.5% of women). The addition of MRI more than doubled the cancer

detection rate of mammography plus ultrasound, but was associated with even an even higher recall rate and a doubling of the biopsy rate. The PPV3 for ultrasound in women with negative mammograms was very low (6.8% round 1, 7.1% rounds 2 and 3) compared to mammography alone (29.1% round one, 38.1% rounds two and three). The PPV3 for MRI in women with negative mammograms was 22.4%.

Summary: Screening HHUS of the Breast

There are no studies evaluating the impact of adding HHUS to mammographic screening among women with dense breast tissue that address the key patient-centered outcomes of breast cancer mortality and disease-free survival. The available body of evidence, focusing largely on shorter-term recall rates, biopsy rates, cancer detection rates and false positive rates, is limited by multiple factors. There were a large number of studies, but the heterogeneity of the study designs, populations, and results preclude the use of meta-analytic techniques to combine the results. The majority of the studies used film mammography, were retrospective, did not fully report the recall rate, and were not able to calculate sensitivity because women with negative mammograms were not followed for interval cancer. There is not even one prospectively designed study with one-year follow-up of HHUS in women with a negative mammogram and heterogeneously dense or extremely dense breasts. The best estimates for sensitivity and specificity come from the ACRIN 6666 trial (87.5% and 81.9% respectively) because it is the highest quality study and sensitivity and specificity are usually not influenced by the risk of the population being studied (Berg, 2012). The best estimate for the incremental cancer detection rate is centered around 3-4 cancers per 1,000 examinations, but the results from the three studies on the Connecticut experience were closer to two cancers per 1,000. The results from Connecticut are more likely to be representative of routine clinical practice in the U.S. The recall rates and PPV1's in these studies were greater than those of mammography, indicating that the addition of HHUS approximately more than doubles the recall rate. The recall rate doubled in the ACRIN 6666 study as well. Finally, the biopsy rates were three to five times higher than those of mammography, suggesting that the biopsy rate of ultrasound after negative mammography is likely to be at least four times that of mammography alone. This is the major limitation of screening ultrasound. The PPV3, which represents the percentage of biopsies that are positive for cancer, was only 7% in studies of women with dense breasts and in the high risk population in the ACRIN 6666 study. The PPV3 in mammography is approximately 25%. Thus the rate of false positive biopsies is much higher with ultrasound. Table ES6 below summarizes the key statistics from the three Connecticut studies (direct evidence) and the ACRIN 6666 study (high quality indirect evidence).

Table ES6: Key findings from the essential, U.S.-based studies of HHUS.

Study	Recall rate per 1,000	Biopsy rate per 1,000	PPV3	Cancer detection rate per 1,000
Hooley 2012	56.7	56.7	5.7%	3.2
Weigert 2012	49.6	48.3	6.7%	3.2
Parris 2013	33.5	32.8	5.5%	1.8
ACRIN 6666	185.7	88.0	6.8%	5.9

PPV3: positive predictive value of biopsies actually performed

The studies comparing mammography, ultrasound, and MRI in very high-risk women described in the section on MRI also help with the comparative effectiveness of the three technologies. In the six studies evaluating all three technologies (Kulh, 2005; Sardanelli, 2007; Warner, 2004; Berg, 2012; Lehman, 2005;

Kuhl, 2010), mammography detected 48 cancers, ultrasound detected 53, and MRI detected 116. Ultrasound detected 19 cancers that were not detected by mammography, which represents 40% (19/48) more cancers detected. Four cancers (3%) were detected only on ultrasound. These studies suggest that the addition of HHUS would increase the cancer detection rate by about 40% (best estimate: 1.4-fold increase) more than mammography alone, but that HHUS does not increase the cancer detection rate when added to mammography plus MRI.

There are no large, well-conducted studies in the United States that directly measure these statistics, which could serve as a reasonable estimate. There is also uncertainty about whether the early detection of these cancers by ultrasound will improve outcomes for women compared to outcomes following their detection as a lump by the women before her next mammogram (interval cancers) or when she has her next screening mammogram.

Estimates based on these data are shown in Table ES7 below. There is a low level of uncertainty around the PPV3 because it was fairly consistent in the literature. The cancer detection rate comes primarily from the three studies describing the experience in Connecticut. There is high uncertainty about the recall rate because of the lack of direct evidence from studies of HHUS in women with dense breast tissue and because of the heterogeneity of the findings in the studies. In all of the prospective studies the recall rate was greater than 100 per 1,000 examinations.

Table ES7: Estimated incremental yield of HHUS after negative digital mammography in women with dense breast tissue.

Statistic	Digital Mammography	Incremental Yield With HHUS	Uncertainty
Recall rate per 1,000	100-120	30-100	High
Biopsy rate per 1,000	14-22	30-60	Low-moderate
CDR per 1,000	3-5	2-4	Low
PPV3	20-25%	5-7%	Low

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

These estimates suggest that HHUS would find 2-4 more cancers than those found by digital mammography alone, with a PPV3 of 6-8%. Recall rates would range widely (30-100) and 30-60 additional biopsies would be required in order to identify these cancers.

Automated Whole Breast Ultrasound (ABUS)

Five studies (Kelly, 2010; Stoblen, 2011; Giuliano, 2013; Arleo; 2014; Brem, 2014) of ABUS evaluating approximately 9,000 participants met the inclusion criteria for the assessment. The primary data are summarized in Table ES8 on the following page; studies are described in detail in the sections that follow.

Table ES8: Key findings from the available studies of ABUS.

Study	Recall rate per 1,000	Biopsy rate per 1,000	PPV3	Cancer detection rate per 1,000
Arleo 2014	188.2	19.7	0	0
Kelly 2010	75.8	12.2	NR	3.8
Brem 2014	150.2	36.0	9.8	7.3
Stoblen 2011	206.9	NR	NR	0
Giuliano 2013	22.8	NR	NR	12.3

PPV3: positive predictive value of biopsies actually performed

Kelly and colleagues (Kelly, 2010) recruited women from eight facilities across the United States. The investigators offered ABUS to consecutive asymptomatic women who had dense breasts. The radiologist reading the mammogram was blinded to the ABUS results and the radiologist reading the ABUS was blinded to the mammography results. Women whose compressed breast thickness at mammography was greater than 7 cm were excluded because of the limited sensitivity of ultrasound at that depth. The percentage of patients who agreed to participate at each site varied from 5% to 25%. The investigators performed 6,425 ABUS examinations in 4,419 women. 1,434 of the examinations were diagnostic examinations because the women had a history of prior breast cancer (776/1,434; 54%), breast implants (399/1,434; 28%), or non-localized abnormalities such as diffuse tenderness or nodularity (159/1,434; 11%). One third of the mammograms were digital and two-thirds were film. Some women at high risk preferred to alternate mammography and ABUS examinations at six-month intervals. The study followed women for one year for interval cancers. The percentage with complete follow-up was not reported, but 5,089 of the women (80%) had a repeat mammogram at least one year after the original mammogram.

The ABUS examination took five to 10 minutes preparation time and 10 to 20 minutes for the examination (Kelly, 2010). The interpretation and reporting time for the radiologist was seven to 10 minutes. The study sample had a median age of 53 years, but included women as young as age 24 and as old as 89 years. The sample included women with a personal history of breast cancer (10%), at least one first-degree relative with breast cancer (30%), and at least one second-degree relative with breast cancer (29%). These proportions are higher than in a typical screening population, suggesting that women who enrolled in the study were at higher risk for breast cancer than the general population.

During the study 23 breast cancers were detected with mammography, 23 by ABUS in women with negative mammograms, and an additional 11 presented as interval cancers that were not detected by either modality. One woman was diagnosed with bilateral breast cancer (56 participants diagnosed with breast cancer). The results are not presented separately for women with negative mammograms, but the statistics for ABUS after a negative mammogram can be calculated from the data presented in the tables and the results section (see Appendix B). The sensitivity of mammography plus ABUS was higher than that of mammography alone (67.6% compared to 41.1%), but the recall rate for mammography plus ABUS was almost double that of mammography alone (74.8 per 1,000 compared to 32.4 per 1,000).

The biopsy rate was also higher with mammography plus ABUS (12.2 compared to 9.1 per 1,000). The cancer detection rates were similar (3.8 compared to 3.6 per 1,000).

There are many methodological concerns that limit the ability to generalize the results of this study to women with dense breasts and a negative digital mammogram. The low volunteer rate (5%-25%) in this study raises concerns about spectrum bias – those who agree to participate may differ from those who do not participate in ways that impact the study results. For instance, women at higher risk for breast cancer may be more likely to volunteer for this study of additional imaging. The wide age range (down to age 27 years) also suggests that this was not a typical screening population. Two-thirds of the mammograms were film, which has much lower sensitivity than digital mammography in women with dense breasts – some of the cancers identified on ABUS would have been picked up by digital mammography (Kerlikowske, 2011; Pisano, 2005). In addition, some of the women elected to be screened with ABUS six months after the mammogram, so the cancers identified by ABUS may represent a mix of interval cancers and those missed by mammography. All of these biases would tend to increase the cancer yield of ABUS.

In the second, much smaller study (Stoblen, 2011), Stoblen and colleagues described the results of ABUS in 304 consecutive women between the ages of 50 and 69 who were seen for routine screening mammography in Germany. The majority of the women had non-dense breasts (scattered fibroglandular densities). All subjects had digital mammography followed by ABUS. Two cases of DCIS were detected by mammography, neither of which was detected by ultrasound. The investigators reported 60 false-positive assessments by ABUS (20.7% of negative mammograms) compared to 12 (4.0%) for digital mammography in the same women. Thus the false positive rate for ABUS was 207 per 1,000 examinations compared to 40 per 1,000 for mammography. However, it does not appear that all of the positive ultrasound findings were biopsied. In addition, no follow-up was reported other than for two patients with repeat examinations at six months, with no additional cancers identified. The study is small and does not directly apply to women with dense breasts, but it highlights the concern about high numbers of false positive results with either automated or hand held ultrasound.

Giuliano and Giuliano (2013) report on the performance of digital mammography plus ABUS in 3,418 asymptomatic U.S. women with dense breasts, compared to 4,076 asymptomatic women with dense breasts screened with digital mammography in the prior year. It is unclear if consecutive women were included. The BI-RADS categories were not used to define high density, but it is likely that the women studied (mammograms with “a Wolfe classification of 50% or greater”) were similar to the two high-density BI-RADS groups. The study excluded women with major risk factors for breast cancer including those with a personal or family history of breast cancer and those with a BRCA mutation. The study was performed at a single site in Florida. Two radiologists read each of mammograms and the ABUS images with final readings by consensus. There was no blinding of the radiologists, but the investigators blinded the pathologists evaluating biopsy specimens.

In the control group, the sensitivity and specificity of digital mammography alone were 76.0% and 98.2% (Giuliano, 2013). The recall rate was 22.8 per 1,000 examinations and the cancer detection rate was 4.7 per 1,000. The biopsy rate was not reported. This cancer detection rate is relatively high for invasive cancer (no cases of DCIS were reported) in women with no personal or family history of breast cancer. The PPV reported (20.4%) is quite high for mammography, suggesting that it may be the PPV for biopsy rather than the PPV for a positive mammography assessment. The low recall rate also supports the under-reporting of recalls for positive results. In the mammography plus ABUS group, the sensitivity was 97.7% and the specificity was 99.7%. Again, the recall rate and biopsy rate are not clearly reported and both are calculated at 15.2 per 1,000 examinations, suggesting that this is actually the biopsy rate.

It is likely that the true recall rate was much higher and that the specificity and PPV are much lower than reported in the paper.

There are several other concerns about this study that call into question all of its results. First, there were no reported cases of DCIS. In 2011, when the study was conducted, approximately 27% of all breast cancer diagnoses were DCIS (DeSantis, 2011). Since the study reported 68 invasive breast cancers, there should have been an additional 25 cases of DCIS. Mammography is more sensitive than ultrasound for the detection of DCIS (Kelly, 2011), so the exclusion of DCIS from the results could have a large impact on the results. It is also worrisome that the results for mammography alone were not reported for the cohort of women also examined with ABUS. It may be that digital mammography performed better in that group because the radiologists had one more year of experience with this relatively new technology. It is also remarkable that the specificity of ABUS was so high. All other reports of ultrasound consistently find a high rate of false positive studies with ultrasound with PPV1 and PPV3 being consistently lower than that of mammography. The opposite was reported in this study. The demographic characteristics of the two groups were not presented, nor compared – if they were very different, then there should have been some adjustment for these differences. The results suggest that there were large differences: the average age for detected invasive cancers in the control group was 54 years, while in the ABUS group it was 57. If ABUS identifies cancers earlier than mammography, then the average age of detected cancers should go down, not up. Given these major concerns, as well as the non-standard breast density measurements and the lack of reporting of the results of ABUS among the women with normal mammograms, the results of this study are not particularly useful in evaluating the appropriate role for ABUS in women with dense breasts.

A small, single-center retrospective study (Arleo, 2014) evaluated 558 ABUS examinations over three months (August-October 2013). The aim of the study was to evaluate differences in the recall rate after implementation of an ABUS system in women with dense breasts (BI-RADS category 3 or 4); if women had been screened previously at this center, results were required to be normal or benign. Women presenting with symptoms or with prior positive ultrasound were screened with HHUS instead. Of the 558 examinations performed on 558 women, 10 women had a family history, three had implants, and 15 had a history of benign breast biopsies. A total of 105 women were recalled for additional screening with HHUS (18.8 recalls per 1,000). The authors noted that this rate improved substantially over the course of three months – from 24.7 per 1,000 in August to 12.6 per 1,000 in October – suggesting an initial learning curve for the new technology, and potential to improve with experience. Eleven women were biopsied, all with benign findings (i.e., PPV3 of 0%). Because of the study's short duration, no other performance characteristics were calculated.

There are some methodological issues with this study beyond the small sample size and short duration. First, because the study evaluated women during the initial implementation of ABUS, the capability of ABUS to reduce the recall rate cannot be fully understood from this analysis alone. Perhaps more importantly, however, there are no data presented on how many women had a prior normal or benign screening result, and how much time had elapsed between that screen and ABUS. It is therefore impossible to put the recall rate presented here in its proper context.

Finally, a very recent prospective multinational study (Brem, 2014) evaluated the use of ABUS plus mammography compared to digital mammography alone in 15,318 women and found that adjunctive ABUS yielded a cancer detection rate of 7.3 cancers per 1,000 women screened compared with 5.4 cancers for screening mammography alone ($p<.001$). However, the recall rate increased from 150.2 recalls per 1,000 women screened with mammography to 284.9 recalls for the combined approach ($p<.001$). PPV1 decreased when adding ABUS (2.6% vs. 3.6% with mammography alone), and required

an additional 36 biopsies per 1,000 women screened. Sensitivity improved significantly by adding ABUS (100% vs. 73.2% for mammography alone), but specificity was reduced (72.0% vs. 85.4%) (p<.001 for both comparisons). Although this study included adequate follow-up and reported results for nearly all outcomes of interest, the initial mammography screenings were not performed by radiologists, which may have introduced some bias by over-assigning women into dense breast categories on the subsequent reading.

Summary: Screening ABUS

None of the studies directly address the use of ABUS following negative digital mammography in a screening population of women with dense breasts. Four of the studies are of poor quality, and only the recent Brem study is of fair quality. Some of these studies (e.g., Kelly, 2010; Brem, 2014) offer reasonable estimates and relatively complete reporting of outcomes, but are of limited applicability because of technical concerns (e.g., use of film mammography, long gaps between mammography and ABUS, use of research coordinators and technologists to assign breast density categories). Across the five studies, the recall rate varied from 5 to 285 per 1,000 examinations, the biopsy rate was not reported or up to 15 per 1,000 examinations, the PPV3 from not reported to 31% and the cancer detection rate ranged from 0 to 7.6 per 1,000 examinations. Overall, the paucity of studies, the lack of high quality studies, and the wide range of estimates across the studies mean that there is considerable uncertainty surrounding all of the estimates for the diagnostic test statistics for ABUS.

Because of the uncertainty described above, we felt that the most reliable estimates for the test characteristics for ABUS come from the HHUS literature, but with high uncertainty. Estimates based on these data are shown in Table ES9 below and are identical to those for HHUS.

Table ES9: Estimated incremental yield of ABUS after negative digital mammography in women with dense breast tissue.

Statistic	Digital Mammography	Incremental Yield With ABUS	Uncertainty
Recall rate per 1,000	100-120	30-100	High
Biopsy rate per 1,000	14-22	30-60	High
CDR per 1,000	3-5	2-4	High
PPV3	20-25%	5-7%	High

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

These estimates suggest that ABUS would find 2-4 more cancers than those found by digital mammography alone with a PPV3 of 5-7%. Recall rates would range widely (30-100) and 30-60 additional biopsies would be required in order to identify these cancers.

Key Question #3: What are the documented and potential harms associated with these imaging tests, including overdiagnosis and overtreatment, unnecessary biopsy as a result of false-positive imaging, patient anxiety, and radiation exposure?

Potential Harms

Overdiagnosis and Overtreatment

An important harm of screening is overdiagnosis: the diagnosis of breast cancers with mammography that, if they had been left undetected, would not have caused symptoms before the woman died of other causes (Zahl, 2008; Welch, 2010). Such patients would endure the toxicity associated with overtreatment of breast cancer (surgery, radiation, hormonal therapy, and chemotherapy), without receiving any benefit of reduced symptoms or longer life from treating the cancer.

None of the studies in our review attempted to measure rates of overdiagnosis or overtreatment based on test performance; for that matter, no long-term clinical outcomes of interest were tracked. This is largely because it is currently impossible to know whether any particular patient whose cancer is detected by mammography is or is not at risk of the cancer being overdiagnosed, and the true magnitude of "overdiagnosis" and subsequent "overtreatment" for breast cancer is unclear and controversial. The most common estimates range from 10% to 30% of cancer diagnoses, although estimates range from as low as 0% to as high as 54% (Blyer, 2005; Welch, 2009a; Jorgensen, 2009; Welch, 2009b; Morrell, 2010; de Gelder, 2011; Wu, 2011; Elmore, 2012; Kalager, 2012; Gotzsche, 2013). This is an area of active research and debate.

Unnecessary Biopsy

The most common harm associated with mammography is a false-positive test result. Approximately 10% of women have a false-positive result at each round of mammography screening and about 50% of women will have at least one false-positive result after 10 mammograms (Rosenberg, 2006; Christiansen, 2000; Elmore, 1998; Hofvind, 2004; Olivotto, 1998; Hubbard, 2011). False positives also usually require that a woman schedule a second appointment for additional imaging resulting in time lost with family or at work and the additional evaluation increases health care costs. Most false-positive results do not lead to a breast biopsy; between 7% and 19% of women have a false positive biopsy after 10 mammograms (Elmore, 1998; Hofvind, 2004).

With regard to the tests of interest in this analysis, biopsy rates ranged from a low of 12 per 1,000 for DBT to up to 60 per 1,000 for ultrasound. Because the overall risk of breast cancer in any large screening population is relatively low, most of these will be unnecessary (i.e., false-positives); for supplemental tests, this will only add to the burden of false-positives already produced by mammography. While false-positive biopsies have effects on patient anxiety (see below), they are also not without clinical consequence. A recent AHRQ review (Dahabreh, 2014) found that severe complications were rare for all forms of breast biopsy (<1%), but that harms were also generally underreported across all studies. Complications can include local reactions, bleeding, and infection or abscess; there have also been case reports of tumor formation at biopsy sites.

Patient Anxiety

While not well-documented in studies of imaging tests, many patients experience short-term increases in anxiety and psychological stress as a result of being recalled after an abnormal finding (Barton, 2001; Barton, 2004; Lipkus, 2000; Scaf-Klomp, 1997; Brett, 2005; Weil, 1997; Woodward, 2001). A systematic review of 23 studies on the long-term effects of false positive mammograms found small, but significant negative impacts on health behaviors and psychological well-being (Brewer, 2007). A recently published case series (Miller, 2013) evaluated women undergoing image-guided breast biopsy and found that stronger physician-patient communication reduced levels of patient anxiety and improved health-related quality of life. Recent advances in online support tools have also shown that increasing interaction between patients and clinicians eases anxiety in women who experience an abnormal screening (Obadina, 2014). Since no imaging modality can be 100% accurate, it is important to consider the use of these anxiety-reducing interventions when weighing the harms and benefits of screening tests.

Radiation Exposure

Ionizing radiation, like that used in mammography or DBT, can damage DNA leading to mutations that increase the risk for the development of cancer. Evidence from those exposed to radiation from the atomic bomb explosions in Japan and from those exposed to radiation therapy as part of treatment for Hodgkin's disease demonstrates that radiation exposure increases the risk for breast cancer (Boice, 2001, Carmichael, 2003; de Gelder, 2011; Ng, 2009; Ronckers, 2005; Yaffe, 2011). The risk is greatest for younger women and is thought to be minimal for post-menopausal women. The radiation dose from mammography is relatively small. The dose from 20 mammograms is equivalent to about three years of environmental exposure to radiation; the dose from one CT scan is equivalent to about 800 mammograms. There is no direct evidence demonstrating an increase in breast cancer due to mammography. One recent modeling study by Yaffe and colleagues estimated that among 100,000 women screened with mammography every year from ages 40 to 55 years and then every two years until age 75 (20 mammograms), the radiation would cause 86 new breast cancer diagnoses and 11 deaths from breast cancer (Yaffe, 2011). Thus for every 1,000 women screened 20 times between the ages of 40 and 75 years, the radiation from mammography may cause 0.9 additional breast cancers and 0.1 additional deaths from breast cancer.

The average dose of radiation from mammography has declined with the transition to digital mammography. In the DMIST trial, the average radiation dose was 4.7 mGy with film mammography and 3.7 mGy with digital mammography (Hendrick, 2010). The Yaffe model assumed that the dose per mammogram was 3.7 mGy based on the DMIST findings (Yaffe, 2011; Hendrick, 2010). Other models using different inputs and assumptions have estimated higher rates of radiation-induced breast cancer and death from mammography (Berrington, 2005).

Although some concerns about the additional exposure to ionizing radiation (double the amount of digital mammography alone) with DBT were initially raised, recent developments in computer software designed to work with existing DBT systems to produce synthesized 2D images from advanced 3D acquisition data have abrogated the need for screening with digital mammography and reduced the radiation exposure to the patient by 50% (Zuley, 2014; Skaane, 2014). The FDA approved the use of such software in May 2013 so evaluation its impact on cancer-related outcomes has not yet been widely studied; nevertheless, once this approach is standardized, it is likely that radiation exposure from DBT will be comparable to that of digital mammography.

Key Question #4: What is the differential effectiveness and safety of the tests of interest according to such factors as age, race or ethnicity, comorbidities, BMI, method of breast density classification, overall breast cancer risk, scan vendor, and imaging protocol (e.g., whether ultrasound is performed by a radiologist, technologist, or some combination of the two)?

Limited data exist to support the use of imaging tests in specific subpopulations beyond the ones included in this assessment (i.e., women with dense breasts or high risk populations). The most common subgroup analyzed in the literature on DBT was age. Three studies (Rose, 2013; Haas, 2013; Destounis, 2014) included subgroup analyses on breast density and age, and both reported similar improvements with DBT for the evaluated outcomes across age groups. One of these studies (Rose, 2013) also evaluated a subgroup of women undergoing screening for the first time, as well as the impact of radiologist experience; however, no material differences in study outcomes were observed. Although no DBT studies evaluated performance across multiple vendors, an editorial observed that the most recent study included in this assessment (Friedewald, 2014) did not use the newest DBT technology, and so considerations of its performance are already outdated (Pisano, 2014).

In the studies evaluating automated whole-breast ultrasound, differential effectiveness related to technologist experience was assessed. Because ultrasonography requires a high degree of technical and medical proficiency, analyzing the imaging protocol of this procedure is important in these early observations (Stoblen, 2011). In a single-center retrospective study (Arleo, 2013) evaluating the first calendar quarter after implementation of ABUS, recall rate improved from 24.7% in the first month to 12.6% at the end of the third. Because ABUS can be operated independently, the investigators suggested an advantage with ABUS over HHUS, particularly with additional experience over time (Arleo, 2013). Moreover, because the automated procedure eliminates operator variability, whole-breast ultrasonography may be more easily reproducible and operators are more confident in their ability to recommend further screening (Kelly, 2010). As noted in the body of the evidence review, however, currently-available ABUS data are too limited to make even indirect comparisons to the other modalities of interest.

One additional study, a retrospective examination of HHUS in Korea (Chae, 2013) examined the impact of age on test performance. Sensitivity was comparable for mammography+HHUS versus mammography alone in all age groups except women age 40-49 (100% versus 29%, $p<.05$). It is difficult to generalize these results to U.S. settings, however, as (a) the overall incidence of breast cancer in Asian countries is less than that in the U.S.; and (b) in contrast to the U.S., where breast cancer incidence increases with age, incidence in Korea is highest among women in their 40s (Chae, 2013).

Key Question #5: What are the costs and cost effectiveness (e.g., cost per cancer detected) of the imaging modalities of interest?

Prior Published Evidence on Costs and Cost-Effectiveness

DBT vs. Digital Mammography

A single, recently-published study examined the cost-effectiveness of screening with DBT vs. digital mammography (Lee, 2014). A discrete-event breast cancer simulation model developed as part of the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET) was used to examine the outcomes and costs of biennial screening with the combination of DBT and digital mammography vs. digital mammography alone among women aged 50-74 years with heterogeneously or extremely dense breast tissue. The test performance of DBT was assumed to be moderately improved over digital mammography (sensitivity 80%, specificity 92% vs. 77%/88% for mammography). DBT screening costs were estimated to be $50 higher for DBT than for digital mammography alone. On a lifetime basis, the use of DBT resulted in 0.5 fewer deaths and 405 fewer false-positives per 1,000 women screened after 12 rounds of screening. Discounted costs were approximately $350 higher for DBT, which delivered an incremental 0.007 QALYs (approximately 3 days) and a cost-effectiveness ratio of approximately $54,000 per QALY gained. Findings were relatively robust to changes in estimated test performance and the disutility of false-positives as well as the additional cost of DBT—cost-effectiveness remained below $100,000 per QALY gained at incremental costs of up to $87 for DBT.

In addition, other studies have examined different strategies for screening mammography in women with dense breast tissue and are summarized here for additional context. Tosteson and colleagues used a simulation model to evaluate the cost-effectiveness of different screening strategies using digital mammography vs. film mammography for U.S. women aged 40 and older, using data from the DMIST trial and other sources (Tosteson, 2008). A strategy of targeted digital mammography (i.e., either for women age <50 and/or women of any age with dense breasts, film in all other women) was estimated to produce more screen-detected cases of cancer and fewer cancer-related deaths than either an all-film or all-digital strategy. Estimates of cost-effectiveness were $26,500 per quality-adjusted life year (QALY) gained for age-targeted digital mammography vs. all-film mammography and $84,500 per QALY for age- and density-targeted digital vs. all-film. A density-targeted digital strategy focused on the Medicare population (age ≥65) yielded cost-effectiveness estimates ranging from $97,000 - $257,000 per QALY gained vs. all-film, depending on assumptions regarding the test performance of digital vs. film mammography.

A more recent study evaluated the performance of multiple digital mammography screening strategies in a U.S. cohort, based on findings from four distinct simulation models (Stout, 2014). Extending annual screening to women age 40-74 with dense breast tissue (with those with non-dense tissue screened biennially) resulted in cost-effectiveness ratios ranging from ~$60,000 - $260,000 per QALY gained relative to screening all women age 40-74 biennially. Variability in model findings was attributed to relatively small incremental benefits from each screening strategy, coupled with high model sensitivity to assumptions regarding women's preferences for avoiding false-positive results. Consistent with the Tosteson study, annual screening strategies in these models that were not targeted by age and/or breast density were not found to be cost-effective.

However, a third study employed a Markov model to compare biennial and annual screening mammography among women with dense breast tissue in Canada (Pataky, 2014), and found that,

compared with a biennial approach, annual screening produced a very small increase in QALYs (0.0014, or less than one day) with increased costs of over $800 per patient. An annual approach produced a cost-effectiveness ratio of approximately $570,000 CAD ($510,000) per QALY gained vs. biennial screening, and had a 37.5% probability of being cost-effective at a willingness-to-pay threshold of $100,000 per QALY gained.

Supplemental Screening in Women with Dense Breast Tissue

Economic evaluations of supplemental screening strategies have directly assessed costs using primary data collected from cohorts of women undergoing screening. One of these reported actual cancer detection and costs from a series of 5,227 asymptomatic Italian women with dense breast tissue and negative mammograms who had HHUS within one month of film mammography. Costs included those of HHUS, clinical examination, biopsy, and cytology, and totaled €56 ($77) per HHUS-screened woman. HHUS detected two additional cancers in this cohort (0.4 per 1,000), resulting in a cost per additional cancer detected estimate of €146,497 ($200,701). The authors hypothesize that the cancer detection rate observed in this study, which was much lower than that reported in the HHUS studies summarized in this review (range: 1.8 – 14.2 per 1,000), may have been a result of self-selection. The sample was limited to women who presented for HHUS within one month of negative mammography, which represented approximately 20% of all women screened at the study site who had dense breasts and negative mammograms. In addition, 72% of women in the study sample were age <50, which is not reflective of the age distribution of women in the general screening population or of the subset with dense breast tissue.

Data are also available from two of the three cohort studies reporting ultrasound experience following the passage of Connecticut's breast density legislation (Hooley, 2012; Weigert, 2012). Hooley and colleagues estimated the cost of providing breast ultrasound to a cohort of 935 mammographically-negative women with dense breast tissue who were screened with HHUS at Yale-New Haven Hospital after passage of the law (Hooley, 2012; Tosteson, 2008). The incremental cancer detection rate was 3.2 per 1,000 screened. Costs, including those of HHUS, aspiration, and biopsy, totaled approximately $180,000 for the cohort, or $60,000 per additional case of cancer detected.

A larger retrospective study of HHUS screening in nearly 9,000 women with dense breast tissue and negative screened mammograms was conducted in 6 radiology practices in Connecticut for the year after passage of the breast density legislation (Weigert, 2012). The cancer detection rate was also 3.2 per 1,000 screened in this study. Costs were estimated for screening and biopsy based on billed charges to insurers, and totaled approximately $3.1 million, or $110,000 per additional cancer detected. Neither study compared screening costs after passage of the law to costs incurred before the law was passed.

Overview of the Cohort Model

As described above, the published literature on the clinical and economic impact of DBT in any population, and of supplemental breast cancer screening modalities in women with dense breast tissue, is noticeably limited. We therefore developed a cohort model to perform a population-based, one-year analysis of clinical and economic outcomes specific to the state of Washington. In the model we included all women age 40-74 except for those with certain high-risk factors, including genetic susceptibility, personal history of breast cancer, and prior chest radiation. Outcomes and costs included those of screening, diagnostic workup (including biopsy when performed), and detection and workup of interval cancers. Costs of treatment and other measures beyond one year of follow-up were not considered. Imaging and biopsy costs were based on the Medicare fee schedule, including the newly-published additional payment for DBT (approximately $57).

As this review has highlighted, the performance of digital mammography varies according to level of breast density. We first conducted baseline analyses comparing the one-year screening performance and costs for both digital mammography and DBT for all women undergoing screening. Because available DBT studies are both incomplete with respect to measurement of sensitivity and specificity and lacking detail on DBT's performance by category of breast density, we tested various possible levels of improvement in test performance relative to digital mammography in our analyses.

Then, we used the model to compare the performance and costs of supplemental screening with each of the modalities of interest (i.e., HHUS, ABUS, and MRI) for women in BI-RADS density categories "c" or "d" who had an initial negative mammogram. For these analyses of supplemental screening, digital mammography was assumed for initial screening, as evidence indicates it is the current screening standard. However, we also examined the performance of these supplemental modalities in which DBT was the screening standard, and women with a negative DBT would then go on for further screening. As with frontline screening, outcomes and costs of supplemental screening were tracked over one year.

We defined the supplemental screening population as a hypothetical cohort that was stratified into different levels of underlying breast cancer risk. Specifically, we divided risk into three levels (low, moderate, and high) that would be based on the woman's age, breast density, and family history of breast cancer -- information likely to be available through physician-patient discussion in the primary care setting. Several more sophisticated risk assessment algorithms are available, but for modeling purposes we opted to use a simplified risk algorithm based on just these three factors to maximize the feasibility and potential generalizability of this approach (see "Overall Breast Cancer Risk" below).

We had to make several broad assumptions in designing the model that are important because they limit the ability of the model to capture the nuances of patient behavior and the many variations in clinical care patterns that occur for individual patients. For example, we assumed perfect compliance for both mammography/DBT and supplemental screening in this analysis. While it is the case that actual compliance is always less than 100%, differences across studies in the definition of the time interval within which women are considered compliant as well as considerations of what constitutes screening vs. diagnostic mammography (Partin, 1998) precluded our use of a uniform, widely-accepted estimate for compliance across different imaging modalities.

The model also assumes that supplemental screening would occur immediately after a negative mammography or DBT result, and that one year of follow-up is available as the reference standard for both mammography/DBT and supplemental screening results. For mammography and DBT, we needed to estimate as "inputs" several important numbers based on our review of the clinical evidence, including the number of cancers detected (i.e., true positives), cancers missed (i.e., interval cancers), recalls for further testing, biopsies performed, cancer "yield" per biopsy (i.e., percentage of biopsies with positive results) and false-positive results both after biopsy and without biopsy (i.e., recalled for further testing but no biopsy recommended). We developed similar inputs for each supplemental screening strategy, but we made a simplifying assumption that all positive supplemental screening tests would result in immediate biopsy, and so did not estimate recall rates (which would equal biopsy rates in this case) or false-positive results without biopsy. As noted in this review, supplemental screening has the potential to detect both cancers missed by mammography or DBT and additional cancers that would not have presented during the interval between mammography/DBT screenings; we therefore included both types of cancer in our estimates for each supplemental modality.

Overall Breast Cancer Risk

As described above, we limited risk factors for breast cancer in our assignment of risk category to age, breast density, and close family history (at least one 1st degree relative), factors that would be readily available for discussion in a primary care setting. The percentage of women with dense breast tissue *and* a close family history was estimated to be 22.7% based on data from a New Hampshire mammography registry study (Titus-Ernstoff, 2006). Using these three risk factors alone in the BCSC Risk Calculator (Tice, 2008), we defined categories of low, moderate, and high risk as below:

Low: BI-RADS density 3 or 4, age 40-49, no close family history (corresponds to 5-year risks generally <1.7%). Risk assumed in the model: 1% (0.2% per year)

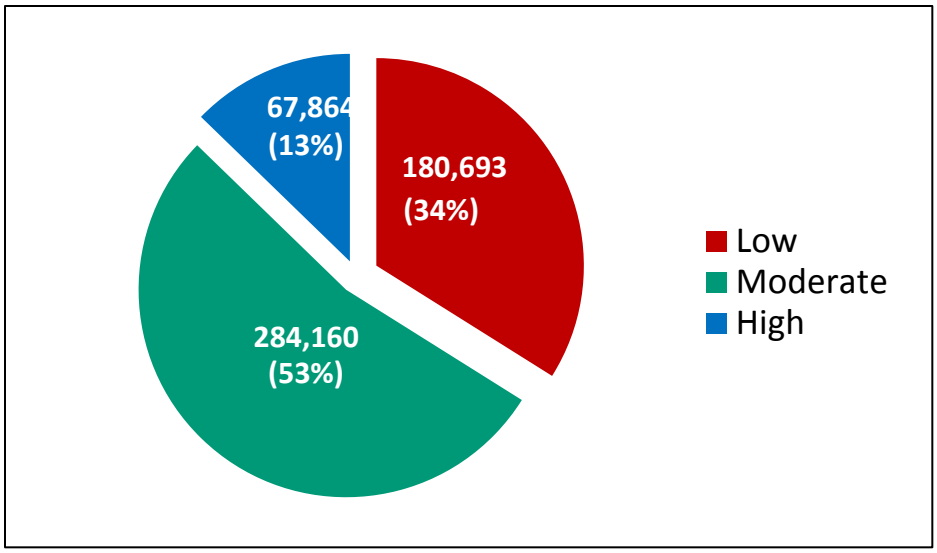
Moderate: BI-RADS density 3 or 4, age 40-49, with a close family history; OR BI-RADS density 3 or 4, age 50-74, no close family history (corresponds to 5-year risks generally between 1.7% and 3.0%). Risk assumed in the model: 2.5% (0.5% per year)

High: BI-RADS density 3 or 4, age 50-74, with a close family history (corresponds to 5-year risks generally >3.0%). Risk assumed in the model: 5.0% (1.0% per year)

Support for these thresholds is available in the literature. Studies of chemoprevention generally consider a 5-year risk of approximately 1.7% to be the lower threshold for considering prophylaxis with tamoxifen or other measures to reduce breast cancer risk (Fisher, 1998; Gail, 1999) while the U.S. Preventive Services Task Force's consideration of the same topic categorized women with 5-year risks >3% to be "higher than average risk" (USPSTF, 2013). Our 5-year risk estimate of 5.0% is comparable to the more commonly-used lifetime risk threshold of >20%, which is often listed as a criterion for MRI supplemental screening in a high-risk cohort.

Based on the risk categories described above, we estimate that, of all Washington women with dense breast tissue and a negative digital mammogram, 34% would be low-risk, 53% moderate-risk and 13% high-risk. These proportions are displayed in Figure ES1 on the following page along with the relevant estimated population sizes for each risk group.

Figure ES1: Estimated numbers of Washington women with dense breast tissue and negative mammography results, by level of overall breast cancer risk.



Model Results

Population Estimates

As mentioned previously, 47% of the 1.3 million women age 40-74 in Washington expected to undergo mammography screening would have BI-RADS density "c" or "d" (620,000). Of these women with dense breasts, 86% (533,000) would be expected to have a negative digital mammogram and therefore be candidates for supplemental screening. Use of DBT as the frontline screening modality would increase the candidate population slightly (to 542,000) as a result of improved specificity – in other words, fewer DBT-screened women would be recalled for further testing and biopsy, and instead be classified as having "negative" or "normal" results.

Comparison of DBT vs. Digital Mammography

The expected performance of DBT vs. digital mammography in Washington is compared in ES10 on the following page for the overall screened population as well as the subset of women with dense breast tissue. To facilitate comparisons, we present all clinical findings on a "per 1,000 women screened" basis, and costs are presented as an average per woman screened.

As shown in Table ES10 on the following page, DBT results in a small increase in the number of cancers detected (3.7 vs. 3.6 per 1,000 for digital) and a small decrease in the number of cancers missed (0.6 vs. 0.7 per 1,000) when compared to digital mammography for the overall screening population. Rates of false-positive results with or without biopsy were both lower for DBT, owing to its slightly better specificity and reduced recall rate relative to digital mammography. However, screening costs were *not* offset by reduced levels of diagnostic workup. Total costs per woman screened were $56 higher for DBT ($245 vs. $189 for digital mammography), meaning that only 2% of the $57 premium in screening cost for DBT would be offset by reduced levels of unnecessary diagnostic workup. Because the estimated costs of recall without biopsy are <15% of the costs of recall with biopsy (i.e., $190 vs. $1,493 respectively), most of the savings from reduced recall are washed out by increased costs of biopsy with DBT.

Not surprisingly, recall and biopsy rates were higher in the subset of women with BI-RADS "c" or "d" breast density, as the incidence of cancer was higher with increasing breast density in the BCSC cohort. For example, cancer occurred at a rate of approximately 5 per 1,000 in women with extremely dense breasts, vs. 2 per 1,000 in those with fatty breasts (BI-RADS 1).

Among women with dense breast tissue, absolute levels of recalls and biopsies performed were higher in comparison to the overall cohort. However, because no density-specific differences in diagnostic performance were assumed for DBT relative to digital mammography, incremental differences were the same as in the overall cohort (i.e., differences of 0.1 per 1,000 in both cancers detected and cancers missed).

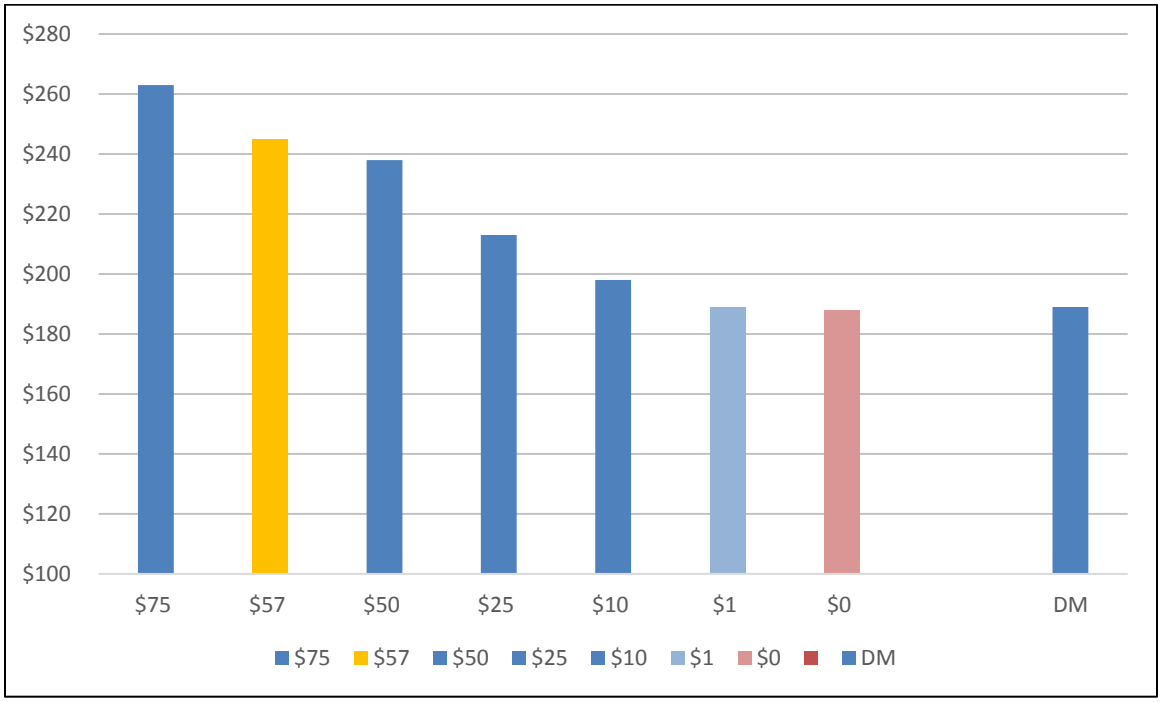
Table ES10: Clinical outcomes and costs of general population breast cancer screening in Washington: comparison of digital mammography vs. digital breast tomosynthesis.

Outcome (per 1,000 screened)	Digital Mammography	DBT
Overall Population		
Recalls	107.0	91.0
Biopsies Performed	18.1	19.3
Cancers Detected (True Positives)	3.6	3.7
False Positive (with Biopsy)	14.5	15.6
False Positive (without Biopsy)	83.3	67.2
Cancers Missed (Interval Cancers)	0.7	0.6
Cost (per Woman Screened, $)	189	245
Women w/Dense Breast Tissue		
Recalls	130.6	114.6
Biopsies Performed	22.1	24.3
Cancers Detected (True Positives)	4.2	4.3
False Positive (with Biopsy)	17.9	20.0
False Positive (without Biopsy)	105.7	89.6
Cancers Missed (Interval Cancers)	0.9	0.8
Cost (per Woman Screened, $)	194	249

NOTES: Recalls refer to positive mammograms or DBTs recalled for additional imaging and/or biopsy; findings may not sum perfectly due to rounding

Findings for our sensitivity analyses are presented in Figure ES2 and Table ES11 on the following pages. The basecase cost for DBT is shown on the yellow bar in Figure ES2. Variations in the dollar premium added for DBT affected only screening costs (i.e., performance statistics remained the same). At a premium of $75, the total cost per woman screened with DBT was $263, or a 39% increase over the cost of digital mammography alone. In contrast, at a small premium of $10, total costs per woman screened were increased by only 5%. A DBT-based screening strategy would be cost neutral to that of digital mammography at a $1 premium (as shown on the light blue bar), and would begin to be cost-saving if the previous Medicare reimbursement approach (i.e., no additional payment for DBT) were to hold (as shown on the light red bar).

Figure ES2. Total cost per woman screened, at different payment premiums for digital breast tomosynthesis over digital mammography.



Results of our analyses varying the test performance of DBT can be found in Table ES11 on the following page. Importantly, the budgetary impact of DBT at a $57 premium remained substantial, even in the most optimistic of scenarios regarding diagnostic accuracy. When DBT was assumed to improve only specificity over digital mammography (scenario A), the only parameters to change from primary analyses related to small changes in the numbers of cancers detected and missed. Costs did not materially change from the basecase. More substantial improvements in both sensitivity and specificity were modeled in scenarios B and C. For example, when DBT was assumed to have 87% sensitivity and 93% specificity (scenario B), the recall rate declined by 21% (71.6 vs. 91.0 per 1,000 in primary analyses), and the biopsy rate also declined substantially (15.2 vs. 19.3 per 1,000). The cost per woman screened was estimated to be $242 under these assumptions, meaning that 7% of the additional $57 in DBT payments would be offset by fewer false-positive recalls and biopsies. Scenario C assumed an 89% sensitivity and 95% specificity for DBT; in this analysis, the recall rate would be reduced by over 40% relative to the basecase (51.7 vs. 91.0 per 1,000); the biopsy rate would be 11 per 1,000 (vs. 19.3 per 1,000 in the basecase). The cost per woman screened would further decline to $238, meaning that 14%

of the additional payments for DBT would be offset by improved test performance in this scenario. It should be noted, however, that a specificity level this high was only reported in the Norwegian study (Skaane, 2013); as discussed, the process of adjudicating breast images is more intense than in the U.S., and 95% specificity may not be achievable here. The final scenario assumed that basecase estimates for sensitivity and specificity (85.5% and 91.5% respectively) would apply, but that DBT would detect an additional 1 cancer per 1,000 screened that would not have been detectable between mammography screenings. The only changes were to the cancer detection rate (4.7 vs. 3.7 per 1,000 in the basecase) and the false-positive biopsy rate (14.6 vs. 15.6 per 1,000). Costs were essentially unchanged in this scenario.

Table ES11. Results of sensitivity analyses varying test performance of digital breast tomosynthesis and number of additional cancers detected vs. digital mammography.

Outcome (per 1,000 screened)	DM Basecase Sn: 84.0 Sp: 90.0	DBT Basecase Sn: 85.5 Sp: 91.5	(A) Sn: 84.0 Sp: 91.5	(B) Sn: 87.0 Sp: 93.0	(C) Sn: 89.0 Sp: 95.0	(D) 1 add'l cancer detected
Overall Population						
Recalls	107.0	91.0	91.0	71.6	51.7	91.0
Biopsies Performed	18.1	19.3	19.3	15.2	11.0	19.3
Cancers Detected	3.6	3.7	3.6	3.8	3.9	4.7
False + (with Biopsy)	14.5	15.6	15.6	11.4	7.1	14.6
False + (w/o Biopsy)	83.3	67.2	67.2	52.3	32.4	67.2
Interval Cancers	0.7	0.6	0.7	0.6	0.5	0.6
Cost (per Woman Screened)	$189	$245	$245	$242	$238	$244

Sn: Sensitivity; Sp: Specificity; DM: Digital mammography; DBT: Digital breast tomosynthesis

NOTES: Recalls refer to positive mammograms or DBTs recalled for additional imaging and/or biopsy; findings may not sum perfectly due to rounding

Incremental Effects of Supplemental Screening in Women with Dense Breast Tissue and a Negative Digital Mammogram

We compared the three supplemental screening scenarios (HHUS, ABUS, MRI, and DBT) to no supplemental screening (i.e., digital mammography alone) on an overall basis as well as separately for low, moderate, and high-risk women. Results are described for each group of interest in the sections that follow.

Overall (All Risk Groups Combined)

Findings for the combined population of low-, moderate-, and high-risk women can be found in ES12 on the following page. As discussed previously, neither recalls nor false-positives without biopsy were estimated for these analyses, as all positive supplemental screening results were assumed to result in biopsy. We present clinical results for HHUS or ABUS together, as equivalent performance was assumed. Costs were assumed to differ, however, and are presented separately at the bottom of the Table ES12 on the following page.

The addition of MRI to digital mammography detects more cancers (6.0 vs. 3.8 for HHUS/ABUS). HHUS/ABUS would nearly quadruple the number of biopsies required over digital mammography alone, while biopsies would increase nearly threefold with MRI. Each of the supplemental modalities would identify nearly all of the cancers missed by mammography. MRI was the more costly strategy ($602), however, due to the higher payment rate for the test itself. Costs were $159 and $243 for HHUS and ABUS respectively.

Table ES12: One-year clinical outcomes and costs of supplemental screening in Washington in all women with dense breast tissue and negative mammography: vs. digital mammography alone.

Outcome (per 1,000 screened)	DM+HHUS/ABUS	DM+MRI	DM Alone
Biopsies Performed	67.9	52.6	22.1
Incremental increase	*45.0*	*30.4*	
Cancers Detected (True Positives	8.0	10.2	4.2
Incremental increase	*3.8*	*6.0*	
Adjusted for potential overdiagnosis (low)	*3.4*	*5.4*	
Adjusted for potential overdiagnosis (high)	*2.6*	*4.2*	
False Positive Biopsy	59.1	42.4	17.9
Incremental increase	*41.2*	*24.5*	
Cancers Missed (Interval Cancers)	0.2	0.1	0.9
Incremental improvement	*(0.7)*	*(0.8)*	
Cost (per Woman Screened, $)	353/437	796	194
Incremental increase	*159/243*	*602*	

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

Higher-Risk Women Only

Outcomes and costs of supplemental screening for women at high risk of breast cancer (5-year risk of 5%) are presented in Table ES13 on the following page (detailed results for low- and moderate-risk women can be found in the full report). Greater than 10% of women would undergo a biopsy after screening with mammography and HHUS; over two-thirds of biopsies would come from the HHUS component of screening. Estimated totals of four and 11 biopsies would be required for MRI and HHUS/ABUS to detect each additional case of cancer. MRI would correctly identify all cancers in high-risk women, although approximately one to three of the approximately 11 cancers identified have the potential to be cases of overdiagnosis. In addition, MRI would miss none of the cancers that would have been missed on mammography, while HHUS/ABUS would miss 0.3 cases per 1,000. Differences in false-positive rates are also magnified in the high-risk population. HHUS would produce a rate of false-positive biopsies more than twice that of MRI (65.0 vs. 31.9 per 1,000 respectively). MRI remained the more costly supplemental test strategy of the three modalities; including costs of mammography, an MRI-based strategy would cost over $800 per woman screened in the high-risk group.

Table ES13: Clinical outcomes and costs of supplemental screening in Washington in women at high overall breast cancer risk with dense breast tissue and negative mammography: vs. digital mammography alone.

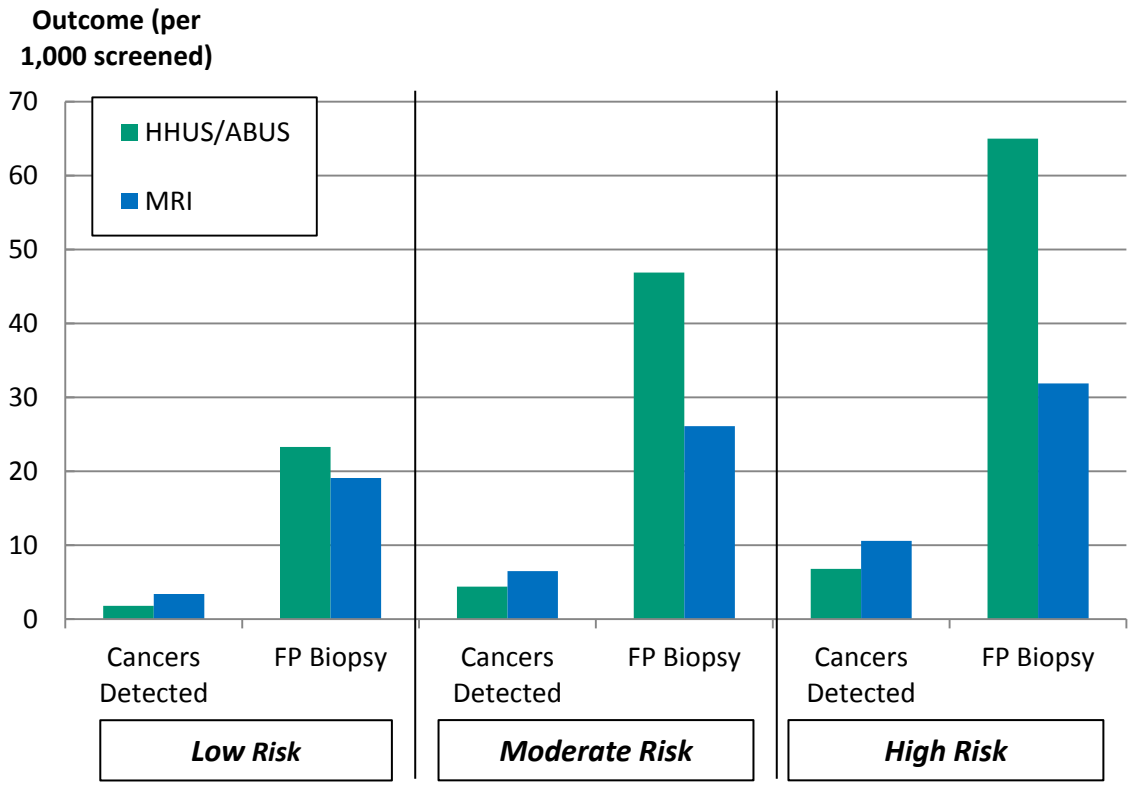
Outcome (per 1,000 screened)	DM+HHUS/ABUS	DM+MRI	DM Alone
Biopsies Performed	110.4	81.2	38.6
Incremental increase	*71.8*	*42.6*	
Cancers Detected (True Positives	14.7	18.5	7.9
Incremental increase	*6.8*	*10.6*	
Adjusted for potential overdiagnosis (low)	*6.1*	*9.5*	
Adjusted for potential overdiagnosis (high)	*4.7*	*7.4*	
False Positive Biopsy	95.7	62.6	30.7
Incremental increase	*65.0*	*31.9*	
Cancers Missed (Interval Cancers)	0.3	---	2.1
Incremental improvement	*(1.8)*	*(2.1)*	
Cost (per Woman Screened, $)	396/480	820	202
Incremental increase	*194/278*	*618*	

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

Comparison of Risk Groups

Key incremental effects of supplemental screening (i.e., above and beyond effects of digital mammography alone) are compared for each risk group in Figure ES3 on the following page. While differences between modalities in the number of additional cancers detected remain relatively stable with increasing risk, differences in rates of false-positive biopsy become more pronounced. For example, HHUS/ABUS would produce 4.2 more false-positive biopsies per 1,000 women screened than MRI in low-risk women (23.3 vs. 19.1 per 1,000 respectively), but would generate 33.1 more per 1,000 among those in the high-risk group (65.0 vs. 31.9 per 1,000).

Figure ES3: Selected incremental effects of supplemental screening, by screening modality and overall breast cancer risk.



HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; FP: False positive

Incremental Effects of Supplemental Screening in Women with Dense Breast Tissue and a Negative Digital Breast Tomosynthesis

We also estimated the effects of supplemental screening with HHUS/ABUS and MRI if frontline screening were performed with DBT rather than digital mammography. Under our basecase analysis, the small improvement in sensitivity with DBT would mean that slightly fewer cancers (0.1 per 1,000) would be missed and available for detection by supplemental screening. The small improvement in specificity would translate into fewer false positives on routine screening, which would have the effect of sending approximately 2% more women with dense breast tissue into supplemental screening. Even with this small increase, however, rates of biopsy and false-positive biopsy would not appreciably change on a per-1,000 screened basis, nor would costs of supplemental screening.

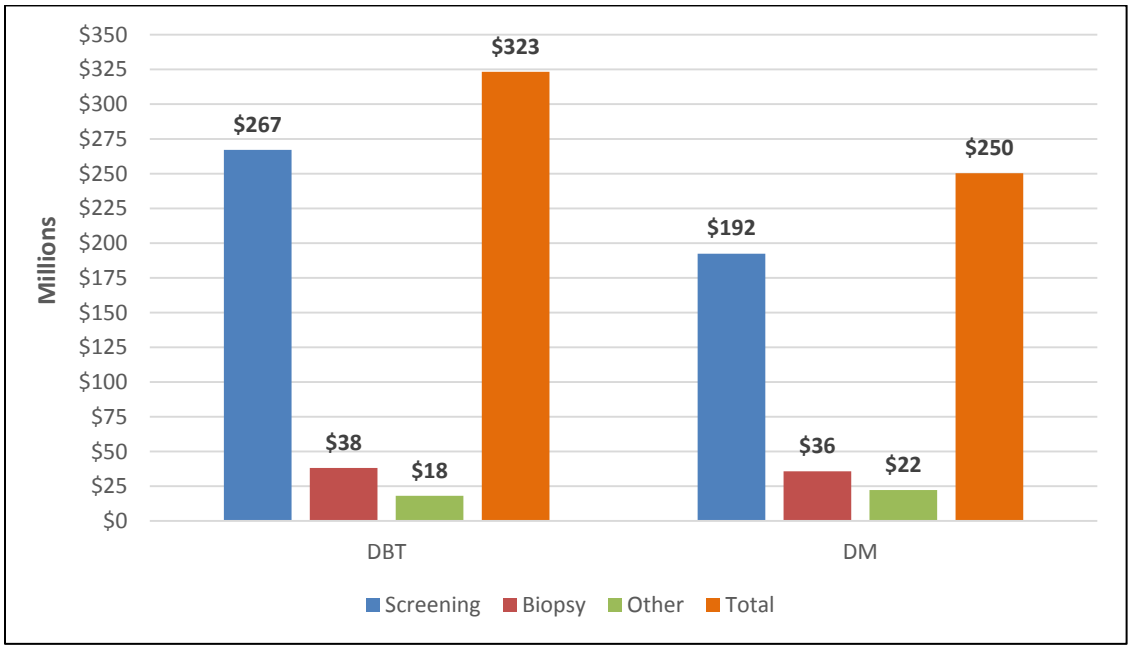
When we assumed that frontline screening with DBT would also detect one additional cancer per 1,000 that would not be detectable by mammography, however, a different picture emerged. Figure ES4 on the following page presents the numbers of cancers detected and false-positive biopsy rates for women at low, moderate, and high overall breast cancer risk. Rates of cancer detection did not change appreciably, as there was only one fewer cancer to detect per 1,000 women screened. However, biopsy

rates (and false-positive rates) declined sharply as a result of DBT's improved performance. For example, among higher-risk women, the false-positive biopsy rate for HHUS/ABUS declined from 65 per 1,000 screened under basecase assumptions to 54 per 1,000 in this scenario. Costs also declined as a result of fewer biopsies. Across all risk groups, the costs per woman screened with HHUS, ABUS, and MRI in this analysis were $142, $226, and $594 respectively; corresponding values in the basecase analysis were $159, $243, and $602.

Population Impact of Frontline and Supplemental Screening

We also estimated the impact of our modeled estimates of frontline and supplemental screening performance and costs when applied to the screening-eligible population in the state of Washington. On a population basis, the costs of frontline screening with digital mammography, including costs of screening, diagnostic workup, and biopsy, are estimated to total approximately $250 million annually. Replacement of digital mammography with DBT in all women would raise costs by nearly 30%, to $323 million annually, using our basecase estimates of +1.5% absolute improvement in both sensitivity and specificity and a $57 premium on screening payment. As shown in Figure ES4 below, savings from reduced recalls with DBT are diminished by higher biopsy costs, so the difference in cost comes almost entirely from the increased cost of the screening test itself.

Figure ES4: Population-based estimates of the cost of routine breast cancer screening among eligible women age 40-74 in the state of Washington.



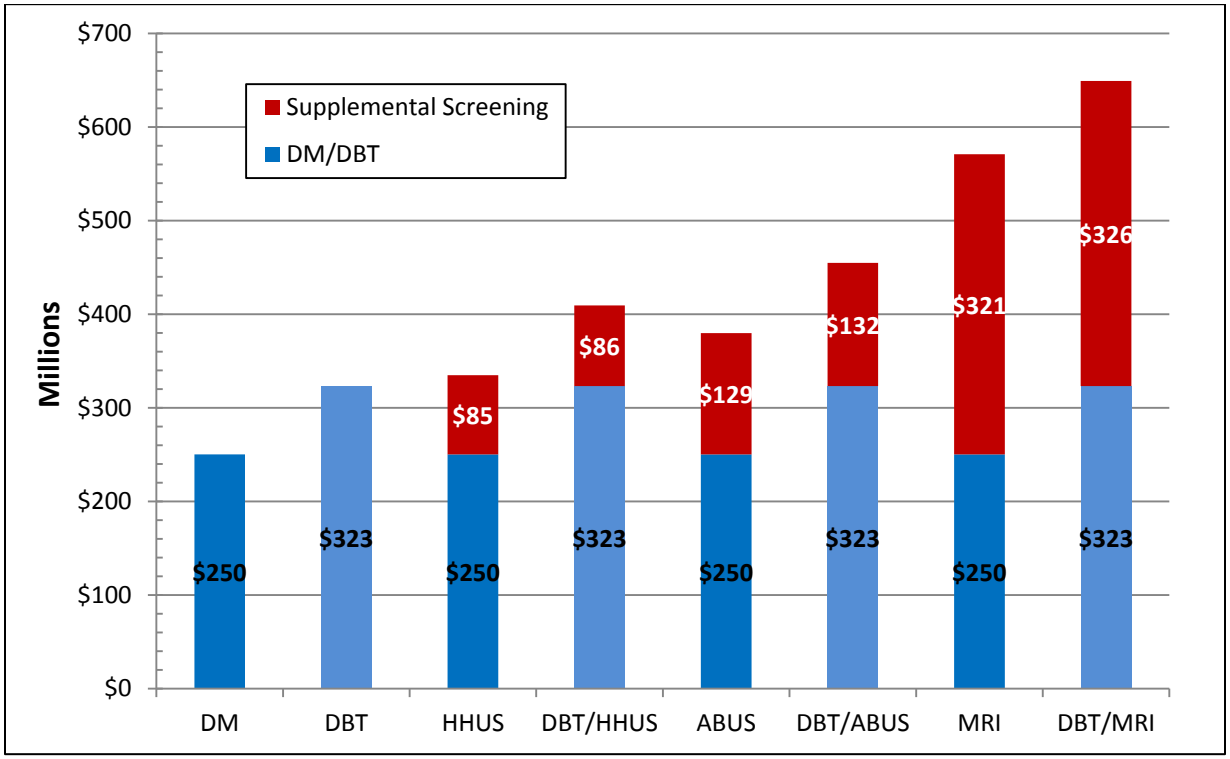
DBT: Digital breast tomosynthesis; DM: Digital mammography

NOTE: "Other" includes cost of diagnostic workup for recalls not involving biopsy and costs of diagnosis for women presenting with interval cancers

The estimated budgetary impact to Washington of supplemental screening in all women with dense breasts and negative mammography can be found in Figure ES5 on the following page. As mentioned above, the annual cost of digital mammography screening is estimated to total approximately $250 million in the state of Washington. Supplemental screening of all women with an initial negative digital

mammogram with HHUS would increase annual costs by approximately 35%, to $335 million. Use of higher-cost ABUS as the modality of choice would result in an increase of over 50% in costs (to $380 million) for the same assumed clinical benefit. Finally, use of MRI results in a more than twofold increase in overall costs (to over $570 million) annually.

Figure ES5: Costs of digital mammography and supplemental screening among all Washington women with dense breast tissue and negative frontline screening result, by screening modality.



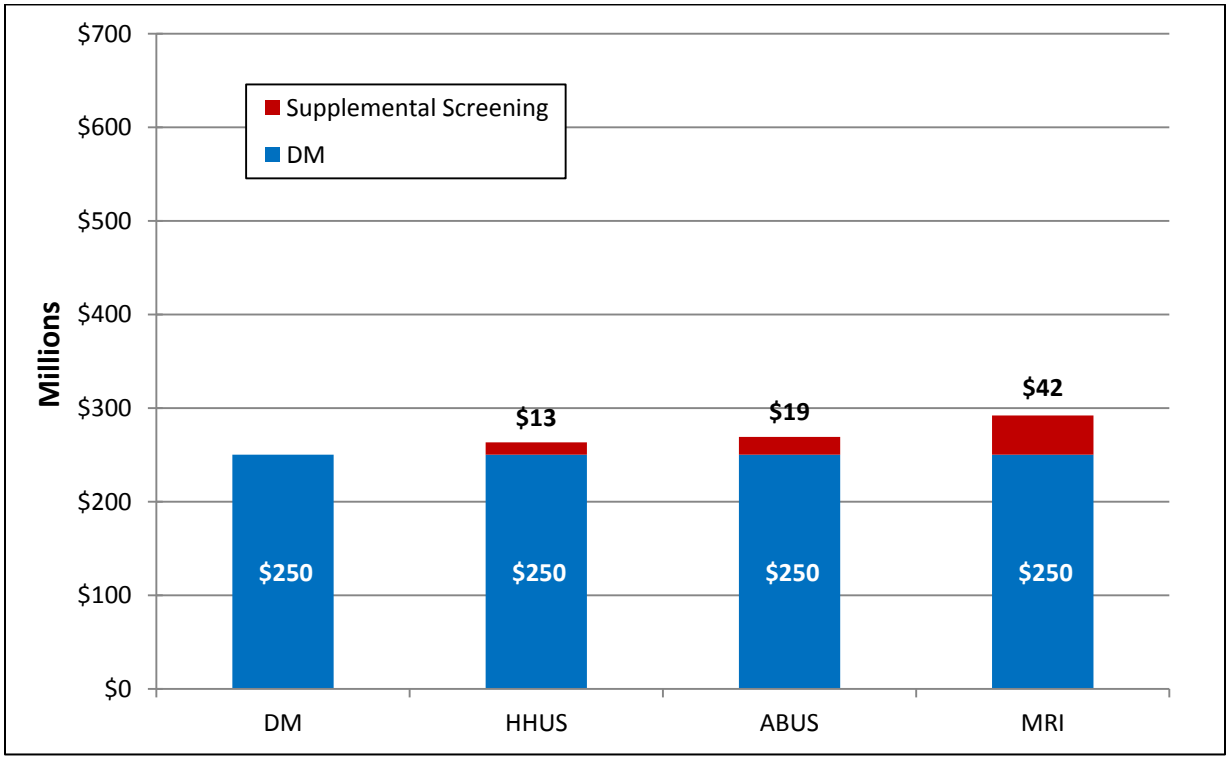
DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; DBT: Digital breast tomosynthesis

When DBT is used as the general population screening modality, total costs of screening are estimated to be $323 million. While most of this increase is due to increased screening costs, costs of supplemental screening also increase slightly (by $1-$6 million depending on the modality used) in this scenario, given the increased size of the population available for supplemental screening. Specifically, approximately 9,000 more women would be screened after a negative DBT; these women would have been recalled for further imaging after digital mammography.

Also, while not represented in the Figure, we examined the impact on total screening costs under the scenario in which DBT detected one additional cancer per 1,000 woman screened that would not have been detectable on mammography. While frontline screening costs did not materially change in this scenario, supplemental screening costs were lower as a result of fewer biopsies required to detect available cancers. Specifically, supplemental screening costs declined by $9 million for HHUS and ABUS, and by $4 million for MRI relative to the primary DBT analysis (i.e., improved sensitivity and specificity, but no additional cancer detection benefit).

Regardless of the frontline screening modality employed, a substantial proportion of the additional costs of supplemental screening are generated in the low-risk population, the subgroup in which the fewest additional cancers are detected. Figure ES6 below shows the additional costs of supplemental screening after negative digital mammography when limited to women in the "higher risk" category.

Figure ES6: Costs of digital mammography and supplemental screening among "*high-risk*" Washington women with dense breast tissue and negative mammography, by screening modality.



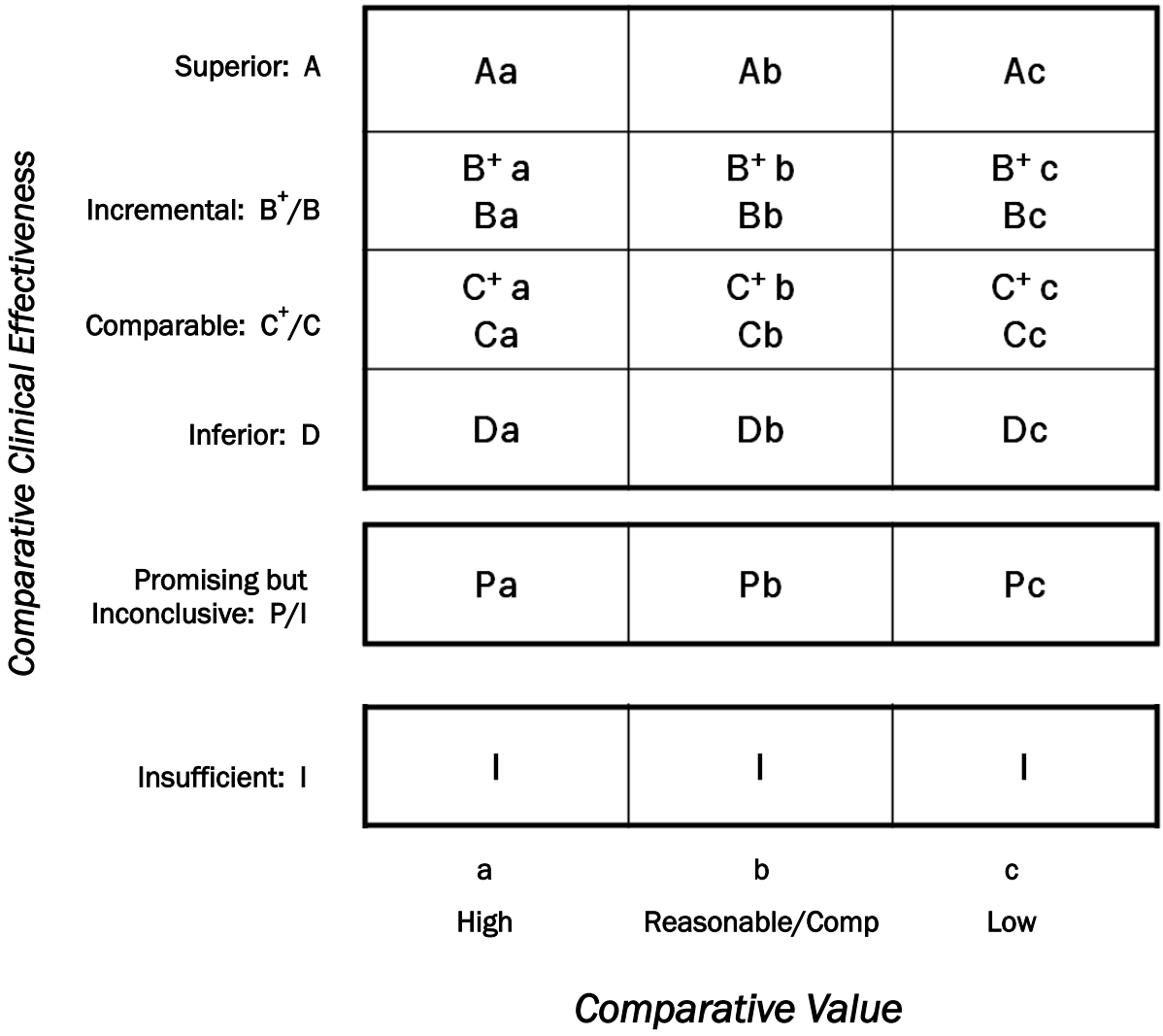
DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; DBT: digital breast tomosynthesis

If supplemental screening were limited to women age 50-74 with dense breast tissue, a family history in a first degree relative, and a negative digital mammogram (i.e., the high-risk cohort), total costs of screening would rise by a much smaller increment. However, the potential yield of additional cancers detected in this subgroup would be comparable to or better than with digital mammography alone. For example, supplemental MRI screening in high-risk women would increase costs by approximately $42 million (17%) to approximately $290 million, and would find a total of 1,358 cases of cancer (636 cancers from digital mammography alone + 722 additional cancers from MRI).

Increases in cost would be lower with HHUS and ABUS (5% and 8% respectively), but the additional cancer yield would also be lower (463 additional cancers detected over digital mammography alone). Findings such as these are important to consider in any evaluation of the tradeoffs of supplemental screening, including numbers of biopsies required, additional cancers detected and missed, and screening costs.

ICER Integrated Evidence Ratings

The ICER integrated evidence rating matrix is shown below; a detailed explanation of the methodology underpinning this rating system can be found in Appendix C to the full report. Separate ratings are provided for each of the screening tests and populations under consideration; the ratings and rationale are described on the following pages. As noted previously, the available literature does not provide guidance regarding the impact of these tests on long-term morbidity and mortality in screening-eligible women. Our focus of attention is therefore on commonly-reported data on test performance, recall rates, biopsy rates, and incremental cancers detected.



Digital Breast Tomosynthesis vs. Digital Mammography for Screening-Eligible Women Age 40-74 Years

1. Comparative Clinical Effectiveness: C+ (Comparable or Better)
2. Comparative Value: Dependent on premium for DBT payment

   a (High) at values <~$30

   b (Reasonable) at values ~$30-$60

   c (Low) at values >~$60

Rationale for ICER Ratings

Of the nine studies identified that compared DBT to digital mammography alone, only a small, single-center study had sufficient follow-up for full determination of test performance (i.e., sensitivity, specificity, negative predictive value, and incremental cancer detection). Nevertheless, while results have been variable, all studies suggest that DBT results in a lower rate of recall compared to digital mammography, has equivalent or better cancer detection, and comparable rates of biopsy. There is no evidence to suggest that DBT's accuracy is in any way inferior to that of digital mammography. While our certainty in these outcomes is moderate, we feel that it is reasonable to assume that longer-term evidence on DBT will show at least comparable (and likely better) performance, leading to our C+ (comparable or better) rating.

There is more uncertainty regarding DBT's economic impact. While CMS has recently published billing codes and payment rates for DBT (~$57), this will not necessarily translate into a uniform premium paid by all insurers. Our model suggests that, using the CMS figures, a very small proportion of increased costs for DBT would be offset by savings in diagnostic workup for recalled patients, as these savings appear to be counterbalanced by higher biopsy costs. It is important to remember that higher biopsy costs are in part due to better cancer detection with DBT, and these costs would likely be at least partially offset by reduced downstream treatment costs as a result of better (and earlier) cancer detection. Indeed, a new publication of a DBT lifetime simulation model suggests that its use may represent a cost-effective use of resources (Lee, 2014); however, this study assumed biennial screening and use of DBT only in women with dense breast tissue.

Under the assumptions of our shorter-term model, the economic impact of DBT is tied more directly to screening costs. This value of DBT would be highest if little or no premium is attached to its use, reasonable at moderate levels of increased payment, and lowest at payment increases of >$60.

Supplemental Screening Among All Women Age 40-74 with Dense Breast Tissue and a Negative Mammogram vs. Digital Mammography Alone

Comparative Clinical Effectiveness:

MRI: B+ (Incremental or Better) [A, Superior]*
HHUS: P (Promising but Inconclusive) [C+, Comparable or Better]*
ABUS: I (Insufficient)

Comparative Value:

MRI: c (Low) [b, Reasonable/Comparable]*
HHUS: c (Low) [b, Reasonable/Comparable]*

ABUS: Not applicable

*Rating in brackets reflects use of supplemental screening in a risk-targeted subgroup

Rationale for ICER Ratings

Ratings of comparative clinical effectiveness for supplemental screening modalities were difficult because there is a lack of evidence in the population of primary interest for this appraisal – i.e., women with dense breast tissue and a normal frontline screening result. MRI, for example, has been studied almost exclusively in women at very high risk for breast cancer. Nevertheless, our view of the evidence suggests that MRI would also detect many additional cancers in women at somewhat lower risk. While MRI's higher sensitivity would also result in recalls and unnecessary biopsies, the high cancer detection rate appears to be an appropriate tradeoff. While our certainty in the evidence is only moderate, MRI's comparative clinical effectiveness relative to digital mammography alone appears to be incremental or better (rating B+). If a screening strategy were developed exclusively for women at higher overall breast cancer risk, however, this is a better match for the current evidence base for MRI, which has demonstrated superior rates of cancer detection relative to no supplemental screening. We would therefore assign a rating of A (superior) in this subgroup.

While the evidence base for HHUS is much more substantial, findings are quite heterogeneous across studies. More importantly, the prospective studies of HHUS suggest rates of recall and false-positive biopsy much higher than those of MRI, and a rate of incremental cancer detection that is much lower. We rated the comparative clinical effectiveness of HHUS to digital mammography alone to be "promising but inconclusive" (rating P) because the benefit/harm balance for this strategy is more unsettled. If the use of HHUS were restricted to high-risk women, we believe the rating would shift to C+ (comparable or better); the evidence base still provides only moderate certainty, but HHUS would have a greater rate of cancer detection in this subgroup. The small evidence base, high degree of variability in findings, and poor study quality led to our "insufficient" rating for ABUS (rating I).

When both MRI and HHUS are used for all women with dense breast tissue, substantial costs are introduced to the healthcare system for the benefits received. On a per-screenee basis, total costs would nearly double with the addition of HHUS to digital mammography, and would be over four times higher through the addition of MRI. We rated the comparative value of both technologies to be low (rating c) in this case because of the substantial budgetary impact. The budgetary impact would be much lower if a risk-targeting strategy such as the one described in our analysis is used. For example, use of MRI exclusively in the higher-risk subgroup would increase the costs of screening in Washington by 17% (vs. an increase in costs from approximately $250 million to $571 million if MRI were used in all women with dense breast tissue and a negative mammogram). We would therefore consider the comparative value of either an MRI- or HHUS-based supplemental screening strategy that is targeted in some way by overall breast cancer risk to be reasonable/comparable (rating b) in comparison to use of digital mammography alone.

We did not rate the comparative value of ABUS because of its insufficient rating for comparative clinical effectiveness.

Appraisal Report

Final Scope

It is estimated that about one in eight women in the United States will develop invasive breast cancer in her lifetime; breast cancer is also the second-leading cause of cancer death among women, behind only lung cancer (BreastCancer.org, 2014). Some women have an elevated risk of breast cancer, including those who have a personal or family history of the disease, genetic abnormalities (particularly carriers of the BRCA1 and BRCA2 gene mutations), previous instances of chest radiation therapy, or the presence of denser, more fibrous breast tissue. These women often undergo supplemental screening with other modalities in addition to conventional mammography, such as ultrasound or magnetic resonance imaging (MRI). In addition, a new technology known as digital breast tomosynthesis is being considered as a potential replacement for conventional mammography among women at all levels of breast cancer risk who are candidates for routine screening.

This review involved an evaluation of the evidence within two distinct constructs: (a) use of digital breast tomosynthesis versus digital mammography as a frontline general population screening tool; and (b) use of automated and handheld ultrasound as well as magnetic resonance imaging for supplemental screening in women with dense breast tissue. This project is an expansion of a previously-conducted systematic review of the published literature on supplemental screening for women with dense breasts (CTAF/CEPAC, 2013). Specific details on the scope of the updated literature search (Population, Intervention, Comparators, and Outcomes, or PICO) are detailed in the following sections.

Objectives and Methods

The objective of this review was to appraise the comparative clinical effectiveness and comparative value of DBT relative to conventional mammography among screening-eligible women, as well as the effectiveness and value of ultrasound and MRI as supplemental screening tests in women with dense breast tissue and a negative mammographic or DBT result. To support this appraisal we have updated the results of a systematic review of published randomized controlled trials, comparative observational studies, and case series on clinical effectiveness and potential harms, as well as any published studies examining the costs and/or cost-effectiveness of these screening strategies. We have also developed a population-based model assessing the potential economic impact of these strategies among screening-eligible women in the state of Washington.

Key Questions

1. What is the effectiveness of screening with digital breast tomosynthesis (DBT) vs. digital mammography among women aged 40-74 who are candidates for screening mammography?
2. What is the comparative effectiveness of handheld ultrasonography, automated ultrasonography, and magnetic resonance imaging when used as supplemental screening modalities in women with dense breast tissue and a negative mammogram or negative DBT result?
3. What are the documented and potential harms associated with these imaging tests, including overdiagnosis and overtreatment, unnecessary biopsy as a result of false-positive imaging, patient anxiety, and radiation exposure?

4. What is the differential effectiveness and safety of the tests of interest according to such factors as age, race or ethnicity, comorbidities, BMI, method of breast density classification, overall breast cancer risk, scan vendor, and imaging protocol (e.g., whether ultrasound is performed by a radiologist, technologist, or some combination of the two)?
5. What are the costs and cost-effectiveness (e.g., cost per cancer detected) of the imaging modalities of interest?

1. Background

1.1 Breast Cancer

Breast cancer is the most common form of cancer in women. An American woman is estimated to have a one in eight chance of developing invasive breast cancer at some time during her life. In 2013, there will be an estimated 234,580 new cases of breast cancer in the United States and an estimated 39,620 deaths from this cancer. This represents approximately 29% of all new cancer cases and 14% of all cancer deaths in women (Siegel, 2013). Moreover, breast cancer is the single leading cause of death for non-smoking women between the ages of 35 and 54 years, accounting for about 10% of all deaths (Woloshin, 2008).

Mortality from breast cancer has declined by about 2.2% per year since 1990, a 28% overall decline (Ries, 2008). The median values from a series of models estimated that a little more than half of the decline was due to improvements in therapy for breast cancer and that a little less than half (46%) was due to early diagnosis from mammography (Berry, 2005). This remains the dominant view, but a recent analysis of 30 years of data from the United States Surveillance, Epidemiology, and End Results (SEER) data called those conclusions into question (Bleyer, 2012). Bleyer and Welch estimated that 31% of breast cancer diagnosed with mammography represents "overdiagnosis" (i.e. , identification of cancers unlikely to cause significant morbidity or mortality) and concluded that screening mammography has had, at best, only a small effect on breast cancer mortality (Bleyer, 2012).

1.2 Screening for Breast Cancer

The primary method used to screen for breast cancer is mammography. Nine large clinical trials established the efficacy of screening mammography by randomizing over 600,000 women and following them for ten to twenty years (Alexander, 1997; Andersson, 1988; Andersson, 1997; Bjurstam, 1997a; Bjurstam, 1997b; Frisell, 1997a; Frisell, 1997b; Miller, 1992a; Miller, 1992b; Miller, 1997; Miller, 2000; Miller, 2002; Nystrom, 2002; Shapiro, 1997; Shapiro, 1988; Tabar, 1995; Tabar, 1989; Tabar, 2000; Moss, 2006). The results have been summarized in many systematic reviews and meta-analyses (Swedish Cancer Society, 1996; Armstrong, 2007; Cox, 1997; Elwood, 1993; Glasziou, 1992; Glasziou, 1995; Gotzsche, 2006; Gotzsche, 2000; Hendrick, 1997; Humphrey, 2002; Kerlikowske, 1997; Kerlikowske, 1995; Nystrom, 1993; Ringash, 2001; Smart, 1995). There is general consensus that, for women between the ages of 50 and 69 years, screening mammography reduces breast cancer mortality by approximately 20% to 25% after 15 years of follow-up (Kerlikowske, 1997). For average-risk women between the ages of 40 to 49 years, there remains significant controversy about whether the benefits of routine mammography outweigh the harms, but most guidelines recommend either routine mammography or a discussion of the benefits and risks of mammography (U.S. Preventive Services Task Force, 2009; Bevers, 2009; Lee, 2010; Smith, 2013).

Digital Mammography

Mammography was traditionally performed with film. It was one of the last radiographic procedures to transition from film to digital imaging because mammography requires extremely high resolution to be effective. Digital image acquisition improves the signal to noise ratio of x-ray detection over a wider contrast range than film (Feig, 1998; Pisano, 2000; Pisano, 1998). Digital enhancement of the images at computer workstations may also improve the accuracy of mammographic interpretation (Pisano, 2000). In particular, increased contrast resolution improves the detection of low contrast lesions in radiographically dense breasts. Digital mammography has become the standard across the United States. As of July 1, 2013, 91.4% (11,705 / 12,800) of all US mammography machines accredited by the

Food and Drug Administration (FDA) are full-field digital (United States Food and Drug Administration, 2013).

Digital Breast Tomosynthesis (DBT)

Digital breast tomosynthesis (DBT) uses a conventional x-ray source that sweeps along an arc around the breast to acquire multiple two-dimensional (2-D) digital images (Houssami, 2013; Kopans, 2013; Sechopoulos, 2013). Breast compression is performed using the same device and technique as conventional mammography. The procedure to obtain each digital view is completed in less than 20 seconds. One of the advantages of DBT is that the images can be acquired immediately following the digital mammogram without needing additional compression. Like MRI, computational algorithms synthesize the resulting 2-D digital images to create tomograms (i.e., slices) allowing for a 3-D reconstruction of the breast. The tomograms can be displayed individually (similar to enhanced conventional mammograms) or in a dynamic movie mode.

The dose of ionizing radiation for DBT is about the same as that used for a conventional mammogram. A standard digital image was acquired in early DBT protocols, which served to increase the total radiation dose to approximately twice that of digital mammography alone (Houssami, 2013; Skaane, 2013). However, software has now been developed to produce a virtual 2-D mammographic image as part of the reconstruction of the tomosynthesis image, which may eliminate the need for a digital mammogram and thereby result in no increased radiation exposure. Early reader studies have shown comparable or better performance with so-called "two-view" DBT in comparison to digital mammography with one-view DBT or digital mammography alone (Zuley, 2014; Rafferty, 2014).

There are other uncertainties with DBT that should be considered in any discussion of its potential to replace digital mammography, however. First, the technology and algorithms used for DBT are still in evolution and have not yet been fully validated (Houssami, 2013, Kopans, 2013; Sechopoulos, 2013). One of the crucial areas is the development of techniques to biopsy lesions that are only seen on DBT (Vialai, 2013), although the FDA recently approved a device (Affirm®, Hologic, Inc.) that can be used with both mammography and DBT systems (FDA, 2013). In addition, the current reading time for DBT is about twice that required for digital mammography (Housammi, 2013; Skaane, 2013). Nevertheless, clinical interest in DBT is very strong. A recent publication summarized a 2012 survey of U.S. breast imaging centers (Hardesty, 2014); in the Western region of the U.S., one-third of centers were offering DBT, a percentage that has almost certainly increased in the two years since the survey.

1.3 Supplemental Screening Modalities for Breast Cancer Screening

There are many imaging approaches to screen for breast cancer in addition to mammography. Magnetic resonance imaging (MRI) has been most widely used. The American Cancer Society first recommended the use of MRI to screen women at highest risk for breast cancer in 2007, based primarily on genetic susceptibility (Saslow, 2007). Hand-held ultrasound has been used as a diagnostic tool to evaluate women with breast masses and has been promoted by some as a screening tool (Mahoney, 2013). The FDA recently has approved automated whole breast ultrasound, which scans and records ultrasound images of the entire breast, for breast cancer screening (Kelly, 2011; Giuliano 2013). Other imaging modalities, such as contrast-enhanced mammography, thermography, diffuse optical tomography, sestamibi, positron emission mammography, dedicated breast computed tomography, electrical impedance scanning, MRI spectroscopy, and breast-specific gamma imaging are still in early investigational phases (Drukteinis, 2013; Yang, 2011) and will not be considered further in this assessment.

All of the advanced imaging technologies considered in this assessment generate multiple two-dimensional images representing slices of the breast. This allows the radiologist to visualize the breast in three-dimensions. This is particularly relevant in mammographically dense breasts because breast cancers may be obscured by superimposed dense tissue.

Magnetic Resonance Imaging (MRI) of the Breast

Magnetic resonance imaging uses strong magnetic fields to image the breast, rather than ionizing radiation. The system uses computational algorithms to generate detailed cross-sectional views of the breast. Mammography requires repositioning of the breast and mammography system for each desired view. In contrast, the MRI examination is typically performed with the patient in the prone position lying on a platform placed in the MR chamber that allows the breast to extend dependently from the patient and does not require repositioning. A contrast agent, gadolinium, is injected through an intravenous catheter (IV) to improve the images of the breast.

In studies of high-risk women, MRI approximately doubles the number of breast cancers that are detected compared to film mammography or breast ultrasound (Berg, 2004; Hagen, 2007; Kriege, 2004; Kuhl, 2005; Leach, 2005; Sardanelli, 2007; Tilanus-Linthorst, 2000; Warner, 2004). However, several factors limit the widespread use of MRI for screening. These include an increase in false positive test results, the need for placement of an intravenous catheter to infuse contrast, the length of time required for the examination, the cost of the examination, limited availability of breast MRI facilities (with special breast-specific magnetic coils and biopsy capability), and contraindications to the use of MRI due to pacemakers and other metallic implants. In addition, mammography has been found to be more sensitive than MRI for the detection of ductal carcinoma in situ (DCIS), a noninvasive cell abnormality in the milk ducts and some invasive breast cancers, so the two are typically used together (Bevers, 2009; Saslow, 2007; Mann, 2008).

Hand-held Ultrasonography (HHUS) of the Breast

HHUS is widely used at breast imaging centers to evaluate breast masses and to guide both cyst aspiration and percutaneous breast biopsy procedures. It is particularly useful to differentiate fluid filled cysts from solid masses (cysts are rarely cancerous). Over time, HHUS has evolved to use higher frequency sound waves to generate images of the breast with improved resolution. In addition, earlier generations of HHUS were not able to penetrate deeply into breast tissue and had a limited field of view. Advantages of ultrasound include the ability to evaluate tissue that is dense on mammography without additional ionizing radiation, which can potentially increase the risk for future cancers. It is also perceived to be more comfortable than mammography because it does not require compression of the breasts.

Ultrasound also has limitations. The primary concern with HHUS is the high number of false positive findings, which often lead to unnecessary biopsies. There are also concerns about the operator dependency and reproducibility of the examinations. Like MRI, HHUS takes time. The average length of time for breast HHUS imaging in a recent study was 19 minutes (Berg, 2008). In that study and many others, a breast radiologist performed the study. At a minimum, the breast radiologist needs to be available to review static images saved by the performing technologist in real time so that additional images can be acquired if necessary.

Automated Whole Breast Ultrasonography (ABUS)

ABUS uses computer driven ultrasound transducers to scan the entire breast under the guidance of a technician. A technician compresses the woman's breasts to her chest wall and applies ultrasound gel. A breast-shaped transducer is placed on the compressed breast and automatically scans the entire breast. The entire procedure, including patient preparation, takes about 15 minutes to complete (Kelly, 2011). ABUS reduces the need for radiologists to perform the scan and decreases the length of time of the exam, thus addressing two of the shortcomings of HHUS. It also produces a scan that should have less operator dependence. The radiologist can review the scan independently using software that displays the images individually or sequentially in a movie mode. The primary drawbacks to ABUS are the inability to image very large breasts, the storage requirements for the data acquired during the scan, and the time required to read the scans (Kelly, 2011).

1.4 Definitions and Statistics used in the Evaluation of Screening Tests for Breast Cancer

In the United States, the Breast Imaging Reporting and Data System (BI-RADS) of the American College of Radiology (American College of Radiology, 2003) sets standards for reporting of the results of breast imaging including mammography, ultrasonography, and MRI. The primary purpose of BI-RADS is to enable consistent reporting and communication regarding findings identified on breast imaging and their management recommendations. In addition widespread BI-RADS use supports quality improvement efforts in breast imaging. There are six standard BI-RADS assessment categories used for women without a known malignancy:

- **0** Incomplete examination, additional imaging or comparison to priors is needed;
- **1** Negative;
- **2** Benign finding(s);
- **3** Probably benign – short interval follow-up suggested;
- **4** Suspicious abnormality – biopsy should be considered; and
- **5** Highly suggestive of malignancy – appropriate action should be taken.

When evaluating screening tests, these results are classified into two categories: a positive test result is any of BI-RADS assessment categories 0, 4 or 5 and a negative test result is any of BI-RADS assessment categories 1, 2, or 3. A true positive is a positive imaging assessment that is followed by a diagnosis of invasive or in situ breast cancer within 12 months. A false positive is a positive imaging result that is not followed by a cancer diagnosis within 12 months. The cancer detection rate is the number of cancers detected by a positive test divided by the number of screening tests performed – for consistency and ease of comparison, we will report it as the number of breast cancers detected per 1,000 screening examinations.

The most common statistics reported by scientists evaluating the diagnostic performance of a test are the sensitivity and specificity. The sensitivity is calculated among women with disease: it is defined as the number of positive tests in women with breast cancer divided by the total number of women with breast cancer and is usually reported as a percentage. In studies of breast imaging, the standard has been to follow women for one year after the screening examination and to count any cancers found during that period as interval cancers. Interval cancers are also known as false negatives because the test was negative, but cancer was likely present. True negatives are the negative test results that remain negative during follow-up. The interval cancers are added to the screen-detected cancers to give

the total number of women with breast cancer for the calculation of these statistics. An important methodological point when assessing studies of diagnostic tests for breast cancer is that if the studies do not follow women with negative test results over time, there will be no way to determine how many of the negative tests missed cancers. When there is no follow-up, there will be no false negative results and the sensitivity will always be 100%.

The specificity of these tests is calculated among women without cancer: it is defined as the number of negative tests in women without breast cancer divided by the total number of women without breast cancer over the 12 month follow-up period and is usually reported as a percentage.

Sensitivity and specificity, while helpful for comparing diagnostic tests, are not that helpful in clinical practice. What clinicians and patients want to know is how likely it is that the patient has cancer if she tests positive and how likely it is that she doesn't have cancer if she tests negative. These concepts are known as the positive predictive value (PPV) and the negative predictive value (NPV). Like sensitivity and specificity, these are usually reported as percentages. The positive predictive value is the number of true positives divided by the total number of positive tests, or the percent chance that a woman with a positive test actually has cancer. The negative predictive value is the number of true negatives divided by the total number of negative tests, or the percent chance that a woman with a negative test does not have cancer. In breast cancer screening, things are more complicated because not every woman with a positive test undergoes a biopsy. The BI-RADS audit of mammography outcomes defines three different positive predictive values. The PPV1 is the traditional definition of the number of true positives divided by the total number with a positive result on imaging, and represents the proportion of cancers identified of women recalled from screening for further diagnostic evaluation. The PPV2 is the number of true positives among those recommended for biopsy (BI-RADS 4 or 5 assessment) divided by the total number recommended for biopsy. Finally, the PPV3 is the number of true positives among all those who actually undergo biopsy divided by the total number of biopsies performed. Mammography audits that are required by law to attempt to track all positive tests allow for the calculation of the PPV but not sensitivity, because many sites do not track women over time for interval cancers (Federal Register, 1997; Linver, 1995).

An important methodological point here is that the predictive values are dependent on the prevalence of cancer. When a diagnostic test is evaluated in two populations, one with a high prevalence of cancer and one with a low prevalence of cancer, the PPV will be higher and the NPV will be lower in the population with the higher prevalence of cancer even though the sensitivity and specificity do not change.

Because of this complexity, two other statistics are also useful: the recall rate is the number of women recalled for additional imaging and/or biopsies divided by the total number of women screened and the biopsy rate defined as the total number of women biopsied divided by the total number of women screened. We will report these statistics per 1,000 women screened to allow for comparison across studies and to allow for comparison with the cancer detection rate. Investigators have reported benchmarks for the PPV of film mammography in the United States. They are based on more than two million screening mammograms in over one million women performed between 1996 and 2002. The data come from six registries in the Breast Cancer Surveillance Consortium (BCSC), a prospective study of breast imaging across the United States. The demographics of participants in this study closely match those of the US population in terms of rural/urban mix, race, Hispanic ethnicity, education, and economic status. The study sample included women ages 40-49 years (29%), 50-79 years (62%), as well as women outside this age range (9%). Approximately 6.3% of the women reported a personal history of breast cancer and 15.2% reported a family history of breast cancer (Rosenberg, 2006). The results

described in this analysis did not include follow-up data so the sensitivity, specificity and negative predictive value could not be calculated. The benchmark PPV statistics come from radiologists performing at least 1,000 mammograms over the study period.

Table 1: Performance benchmarks for screening mammography.

Statistic	BCSC Value
PPV1, %	4.8
PPV2, %	25.0
PPV3, %	32.6
Recall rate, per 1,000	94
Biopsy rate, per 1,000	10
Cancer detection rate, per 1,000	4.7
DCIS, %	21.6
Cancers ≤ 10 mm, %	37.2
Node negative, %	79.8
Stage 0 or 1, %	75.6

PPV1 = PPV based on a positive result on initial imaging; PPV2 = PPV based on a recommendation for biopsy; PPV3 = PPV based on biopsies actually performed.

Thus, across the United States, for every 1,000 mammograms performed approximately 100 women will be recalled and 10 will have a biopsy to detect about 5 cancers. One of those cancers will be DCIS (~20%), four will be lymph node negative (~80%), and 3 or 4 (~75%) will be stage 0 or 1 (Rosenberg, 2006). These statistics will vary when looking at different subgroups of women or different screening technologies. For instance, younger women have more false positive mammography assessments and a lower risk for cancer, so their recall rate will be higher and the number of cancers detected will be lower. Digital mammography, which has greater sensitivity and similar specificity compared to film mammography, will have a similar recall rate, but a higher cancer detection rate (Pisano, 2005).

The primary benefit of screening is a reduction in death from breast cancer. As described above, there have been nine large randomized trials evaluating the efficacy of screening mammography (Alexander, 1997; Andersson, 1988; Andersson, 1997; Bjurstam, 1997a; Bjurstam, 1997b; Frisell, 1997a; Frisell, 1997b; Miller, 1992a; Miller, 1992b; Miller, 1997; Miller, 2000; Miller, 2002; Nystrom, 2002; Shapiro, 1997; Shapiro, 1988; Tabar, 1995; Tabar, 1989; Tabar, 2000; Moss, 2006). The studies found that screening mammography reduces breast cancer mortality by approximately 20% to 25% after 15 years of follow-up (Kerlikowske, 1997). In absolute terms, for every 1,000 women screened with mammography for 15 years, there will be 1.8 fewer deaths from breast cancer (Keen, 2009). In addition to the mortality reduction, there may be other benefits, such as less need for aggressive therapies in early stage disease and decreased anxiety about breast cancer following a negative mammogram.

1.5 Mammographic Breast Density

As described previously, mammographic density refers to areas within the breast that absorb significant amounts of x-ray energy and show up as relatively white areas on the mammogram. These correspond to regions in the breast that are rich in epithelial and stromal tissue while the non-dense (darker gray areas) correspond to regions that are predominantly fat.

Breast Density and Masking

In the United States, the Breast Imaging Reporting and Data System (BI-RADS) of the American College of Radiology (American College of Radiology, 2003) classifies density in the following four breast composition categories as listed below:

1. The breasts are almost entirely fatty;
2. There are scattered areas of fibroglandular density;
3. The breasts are heterogeneously dense, which may obscure small masses; and
4. The breasts are extremely dense, which lowers the sensitivity of mammography.

The majority of mammograms in the United States include BI-RADS density as part of the official report (Tice, 2008). Until recently, these BI-RADS categories were further defined quantitatively as less than 25% dense tissue (category 1), 25 to 49% dense tissue (category 2), 50 to 74% dense tissue (category 3), and greater than or equal to 75% dense tissue (category 4). However, the fifth edition of the BI-RADS Atlas, published in late 2013, removes the quantitative descriptors, based on the rationale that a text description of the location and level of dense tissue, and the associated potential to mask cancers and lower the sensitivity of mammography, is clinically more important than the correlation of the percentage of breast density with an increase in breast cancer risk (BI-RADS Atlas, 2013).

It has been known for a long time that the sensitivity of film mammography is lower in women with dense breasts than in women with fatty breasts (Kerlikowske, 1996). There clearly is a masking effect due to mammographic density. In the BCSC registry, the sensitivity of film mammography decreased markedly with increasing density (see Table 2 below). This study evaluated the results from 463,372 screening film mammograms performed between 1996 and 1998. Among women in the lowest density categories, the sensitivity of mammography was 88% and 82% for density categories 1 and 2 respectively, but this decreased to 69% for women with heterogeneously dense breasts and to 62% for women with extremely dense breasts (Carney, 2003).

Table 2: Sensitivity of film and digital mammography by breast density.

		BI-RADS Density Category			
Study	**Type**	**Almost Entirely Fatty**	**Scattered Fibroglandular Densities**	**Heterogeneously Dense**	**Extremely Dense**
BCSC Carney 2003	Film	88.2	82.1	68.9	62.2
DMIST Pisano 2005	Film Digital			55* 70*	
BCSC Kerlikowske 2011	Film Digital	85.7 78.3	85.1 86.6	79.3 82.1	68.1 83.6

*The DMIST study reported results for the combined high-density categories only

Breast Density and Digital Mammography

As described above, the increased contrast resolution of digital mammography improves the detection of low contrast lesions in radiographically dense breasts. Thus digital mammography should improve the sensitivity of mammography in women with dense breast tissue compared to film.

The Digital Mammography Imaging Screening Trial (DMIST) study is the largest trial directly comparing digital mammography to plain film mammography (n=42,760) (Pisano, 2005). All women were screened with both film and digital mammography on the same visit. The mammograms were read independently by radiologists blinded to the results of the other mammogram. In DMIST, digital mammography had the same recall and biopsy rates as film mammography. Digital mammography was more sensitive than film, particularly for younger women with denser breasts (59.1% versus 27.3%, p=0. 0013). Among women of all ages with either heterogeneously dense or extremely dense breasts, digital mammography was also more sensitive than film mammography (70% versus 55%, p= 0.02, Table 2). Similarly, in women with dense breast tissue there was a trend towards greater specificity with digital mammography (91% versus 90%, p=0.09) and the overall accuracy of digital mammography, as measured by the area under the receiver operator curve, was greater than that of film mammography (0.78 versus 0.68, p=0.003) (Pisano, 2008).

The BCSC has recently updated their earlier description of the sensitivity of mammography based on a comparison of 231,034 digital mammograms and 638,252 film mammograms performed between January 1, 2000 and December 31, 2006 (Kerlikowske, 2011). Similar to the prior study, the sensitivity of film mammography decreased from 86% to 68% across the four breast density categories (see Table 2 above). However, for digital mammography, the sensitivity of digital mammography remained greater than 80% for the highest density categories and did *not* appear to decrease with increasing density. As in the DMIST trial, digital mammography was significantly more sensitive than film mammography in women with dense breasts.

Table 3 below shows the cancer detection rate and specificity in addition to the sensitivity of digital mammography by BI-RADS density category in the BCSC study. Despite concerns about the test performance of mammography in dense breasts, more breast cancers are found per 1,000 digital screening mammograms in the denser breast categories than in the less dense categories. This highlights the general principle that the yield of screening tests is greater as the underlying risk of the population screened goes up. Women with denser breasts are at higher risk, so the cancer detection rate is higher. These data also suggest that the masking effect of breast density is minimized when digital mammography is used.

Table 3: Cancer detection, sensitivity, and specificity of digital mammography by breast density.

		BI-RADS density category			
Study	**Type**	**Almost entirely fatty**	**Scattered fibroglandular densities**	**Heterogeneously dense**	**Extremely dense**
BCSC	Rate*	1.8	3.3	4.8	5.1
Kerlikowske	Sens	78	87	82	84
2011	Spec	95	91	87	89

*Rate = breast cancer detection rate per 1,000 women screened

Table 4 on the following page summarizes important outcomes with film and digital mammograms in the three largest studies that report data on both digital and film mammography (Pisano, 2005; Kerlikowske, 2011; Vigeland, 2008). These are useful benchmarks to use when evaluating the potential yield of additional imaging compared to no additional imaging. The biopsy rate and cancer detection rate did not differ between patients screened with digital or film mammography in any of these studies, although the recall rate for digital was higher in the BCSC (100 vs. 93 per 1,000, p<.001). When cancer detection was stratified by breast density in the BCSC, no statistical differences were found between digital and film mammography. However, there was a nominal trend toward higher cancer detection in women with extremely dense breasts (5.1 vs. 3.8 per 1,000, p=.17); the authors concluded that this was primarily due to better detection in women aged 40-49 with extremely dense breast tissue.

It is also worth noting in Table 4 that in Europe, the recall rate for mammography is generally about half that observed in the United States (Hofvind, 2008; Jensen, 2010; Smith-Bindman, 2003). Thus one of the harms of mammography, recalls for false positive imaging results, is less common in Europe. It will be important to keep this in mind when evaluating how to apply the results from studies of supplemental screening performed in Europe to the United States.

Table 4: Recall rates and cancer detection using film and digital mammography in large studies of screening irrespective of density.

Study	Type	Mammograms, n	Recall rate /1,000	Biopsy rate /1,000	Cancer detection /1,000	PPV3
DMIST	Film	42,555	86	16.0	4.1	24.4
Pisano 2005	Digital	42,555	86	15.9	4.4	26.0
Vestfold	Film	324,763	42	NR	6.5	15.1
Vigeland 2008	Digital	18,239	41	NR	7.7	18.5
BCSC	Film	638,252	93	10.6	3.8	24.7
Kerlikowske 2011	Digital	231,034	100	11.0	3.8	25.3

PPV3 = the positive predictive value for biopsies performed

In summary, the findings from both the DMIST and BCSC studies, along with the results from other high-quality studies, highlight a critical difference between digital and film mammography in women with dense breast tissue. The studies find that digital mammography is more sensitive than film mammography in women with dense breast tissue. Therefore the masking effect of breast density observed with film mammography is substantially reduced.

Breast Density and Cancer Risk

The initial report of an association of patterns of mammographic density and breast cancer was published nearly 40 years ago (Wolfe, 1976). The author described four different parenchymal patterns seen on mammography and reported that women with dysplastic pattern with sheets of dense parenchyma had a markedly increased incidence of breast cancer compared to women with normal breast parenchyma. A recent meta-analysis summarizing the literature on the BI-RADS breast density reported a four-fold increased risk for breast cancer in women with extremely dense breasts compared to women with fatty breasts (relative risk [RR] 4.0, 95% CI 3.1 to 5.3), similar to the Wolfe patterns (Cummings, 2009). Risk consistently increased with increasing category of density. Using the more

prevalent group of women with scattered fibroglandular density as the reference group, the risk increases linearly across the four categories (RR 0.5, 1.0 [reference group], 1.5, and 2.0).

If the lifetime risk for breast cancer in the overall population of women is about 12%, then the lifetime risk for women with dense breasts would be approximately 15%. However, lifetime risk is not helpful in deciding when to begin to screen for breast cancer or when to add additional screening – a five or ten year time frame is more clinically relevant. In one study of 629,229 women the observed five-year incidence of invasive breast cancer increased from 7.5 per 1,000 women in the almost entirely fatty group to 12.4 in the scattered fibroglandular density group, 16.5 in the heterogeneously dense group, and 18.1 in the extremely dense group (Tice, 2008).

Because high breast density is both a strong risk factor (relative risk of 1.5 for heterogeneously dense and 2 for extremely dense compared to scattered fibroglandular densities) and it is common (about 40% of women are in the heterogeneously dense category and 10% in the extremely dense category) it explains a greater proportion of the risk for breast cancer in the population than any risk factor other than age. For example, having a first-degree relative with breast cancer almost doubles a woman's risk for breast cancer, but only 10% to 20% of women have a positive family history. Similarly, carrying a BRCA mutation increases a woman's risk by a factor of 10 to 20, but less than 0.5% of women have a deleterious mutation.

One of the common concerns raised about the association between breast density and cancer risk is whether the elevated risk is due solely to the dense tissue masking breast cancers that are present at the time of mammography. If there were only masking, then there would be an increase in cancers detected over the next one to two years in women with dense breasts (those missed on mammography that should have been found) compared to those with fatty breasts, but this excess should not continue beyond two to three years. However, two large studies found no decrease in the strength of the association between breast density and breast cancer incidence through ten years of follow-up (Boyd, 2007; Yaghjyan, 2013). This provides strong evidence that the association of mammographic density with breast cancer represents a true association that is not an artifact arising from the masking of prevalent cancers alone.

Breast Density and Risk Assessment

Risk assessment forms the foundation of all screening and prevention programs. Screening programs using mammography for the early detection of breast cancer generally use age as the primary factor to determine eligibility for screening because age is the strongest risk factor for breast cancer. If the incidence of breast cancer is low, the harms associated with screening outweigh the benefits through early detection and treatment of breast cancer. Conversely, for women at higher risks of breast cancer, earlier and more intensive forms of screening offer the possibility of a more favorable risk-benefit ratio. For example, the American Cancer Society (ACS) guidelines recommend annual MRI screening for women with a lifetime risk for breast cancer above 20% to 25%. This risk threshold was chosen based on expert opinion (Saslow, 2007). The FDA indication for the use of tamoxifen to prevent breast cancer is specifically for women with a 5-year risk greater than 1.66% and, similarly, the 2013 American Society for Clinical Oncology guidelines recommend that physicians consider the use of medications to reduce the risk of breast cancer in women with a 5-year risk greater than 1.66% (Visvanathan, 2013). The 1.66% five-year risk threshold was the primary inclusion criteria for the Breast Cancer Prevention Trial, which demonstrated that tamoxifen reduces the risk of breast cancer by about 50% (Fisher, 1998).

It is worth noting here that using a five or ten-year time frame for estimating risk is more useful than lifetime risk when deciding when to initiate screening for breast cancer. No one would recommend that

a ten year old girl with a lifetime risk for breast cancer of 25% be screened with MRI for breast cancer. Her short-term risk is too low to justify the cost and potential harms. Similarly, a woman with a 2% five year risk for breast cancer by the Gail model could have a 10% lifetime risk or a 30% lifetime risk; in either case she would be eligible for a discussion of the risks and benefits of tamoxifen to lower her risk for breast cancer.

Investigators at the National Cancer Institute developed the most commonly used model of a woman's risk for breast cancer, the Gail model or Breast Cancer Risk Assessment Tool. This model uses a woman's reproductive history and the number of first-degree relatives with breast cancer to estimate her risk for invasive breast cancer (Gail, 1989; Costantino, 1999). A web-based calculator is available for women and their physicians to use: http://www.cancer.gov/bcrisktool/. The model estimates a women's risk of developing invasive breast cancer in the next five years as well as her lifetime risk for invasive breast cancer. The Gail model remains the most widely used tool for estimating a woman's future risk for breast cancer because it was the earliest validated model and it established the entry criteria for the Breast Cancer Prevention Trial.

The limited ability of the Gail model to discriminate high risk women from low risk women has encouraged investigators to develop models that incorporate additional risk factors (Rockhill, 2001). Because breast density is both common and a strong risk factor for breast cancer, researchers have added it to new models (Tice, 2008; Barlow, 2006; Chen, 2006). Investigators at the BCSC developed a model that uses BI-RADS density in combination with a woman's age, race/ethnicity, family history, and history of breast biopsies to estimate her 5-year risk for breast cancer. A web-based calculator using the BCSC model is available for women and their physicians (https://tools.bcsc-scc.org/BC5yearRisk/calculator.htm). The BCSC model has better risk discrimination than the Gail model and is more accurate in non-white women (Tice, 2008). Drs. Chen and Gail updated the Gail model by adding a continuous measure of breast density to their model, but continuous breast density is not routinely calculated or reported with mammography at this time (Chen, 2006).

Mammography screening may be the ideal time for risk assessment because women and their physicians are thinking about breast cancer risk when mammograms are ordered and because mammographic density is the most powerful predictor of breast cancer after age. The new legislation in Connecticut and elsewhere requiring notification of women with dense breasts about the potential for dense breast tissue to mask cancers and increase overall risk makes this an opportune moment for such discussions.

1.6 Washington State Utilization and Cost Data

Figure 1a. PEBB/UMP Breast Cancer Imaging Counts and Costs, 2010-2013

Public Employees Benefits (PEBB)/ Uniform Medical Plan (UMP)		2010	2011	2012	2013	4 Yr Overall **	Avg Annual % Change	
Populations	Average Annual Members	213,487	212,596	212,684	222,339		1.4%	
	Women	117,052	116,313	117,103	121,195		-1.3%	*
	Women 40-74	66,899	66,109	65,965	67,160		-2.3%	*
	Breast Cancer Diagnoses[1]	2675	1065	1013	867	4789	-27.7%	*
	Inconclusive Mammogram [2]	4255	3682	3409	3242	13,164	-9.9%	*
Breast Imaging								
All Breast Imaging	**Total Paid (PEBB Primary)**	**$4,352,472**	**$7,929,564**	**$7,923,481**	**$7,650,299**	**$27,855,816**	**-3.9%**	*
	Procedures (PEBB + Medicare)	47,484	43,629	41,869	42,861	175,843	-3.1%	*
Non digital Mammograms[3]	**Paid (PEBB Primary)**	**$440,748**	**$1,488,224**	**$1,417,310**	**$1,249,133**	**$4,595,415**	**-10.2%**	*
	Average Paid (PEBB Primary)	$61	$150	$174	$171			
	Procedures (PEBB + Medicare)	11,404	13,292	11,352	10,375	46,423	-13.6%	*
	% Screening CPT code (cts)	44.4%	38.0%	33.0%	27.3%			
	% Diagnostic CPT code (cts)	1.1%	1.9%	0.6%	0.1%			
Digital Mammograms[4]	**Paid (PEBB Primary)**	**$3,510,874**	**$5,388,433**	**$5,444,636**	**$5,050,717**	**$19,394,660**	**-5.1%**	*
	Average Paid (PEBB Primary)	$180	$240	$247	$236			
	Procedures (PEBB + Medicare)	32,632	24,197	24,098	24,563	105,490	-1.5%	*
	% Screening CPT code (cts)	99.1%	98.8%	98.8%	99.2%			
	% Diagnostic CPT code (cts)	0.9%	1.2%	1.2%	0.8%			
Tomographic Breast Imaging[5]	**Paid (PEBB Primary)**	**$0**	**$21,603**	**$123,576**	**$453,027**	**$598,206**	361.2%	*
	Average Paid (PEBB Primary)		$257	$253	$261			
	Procedures (PEBB + Medicare)		85	501	1,826	2,412	368.9%	*
US Breast Imaging[6]	**Paid $ US (PEBB Primary)**	**$111,609**	**$478,872**	**$485,464**	**$518,774**	**$1,594,719**	**1.8%**	*
	Average Paid (PEBB Primary)	$37	$99	$104	$107			
	Procedures (PEBB + Medicare)	3,060	5,240	5,098	5,303	18,701	-1.6%	*
Breast MRI[7]	**Paid $ BMRI (PEBB Primary)**	**$289,241**	**$552,432**	**$452,495**	**$378,648**	**$1,672,816**	-19.0%	*
	Average Paid (PEBB Primary)	$771	$720	$584	$512			
	Procedures (PEBB + Medicare)	388	815	820	794	2,817	-3.4%	*

Public Employees Benefits (PEBB)/ Uniform Medical Plan (UMP)	2010	2011	2012	2013	4 Yr Overall **	Avg Annual % Change	
Biopsy (non-imaging comparator)[8]							
Paid $ Bx (PEBB Primary)	**$268,388**	**$1,464,903**	**$1,450,671**	**$1,406,205**	**$4,590,167**	-4.1%	*
Average Paid (PEBB Primary)	$343	$851	$884	$854			
Procedures (PEBB + Medicare)	796	1,783	1,672	1,694	5,945	-4.7%	*

Notes:

Payments for services where UMP is not prime are not reliably reported in the PEBB/UMP claims database, so are excluded from reporting.

Member counts are not included because they are almost the same as the procedure counts. Annual usage rates for most procedures are less than 1.05 procedures per member per year. Exceptions were: biopsy rates approach 1.08 procedures per member per year, and ultrasound and non-specific mammogram rates were as high as 1.16 for Medicare Primary members.

Some PEBB/UMP statistics appear to be widely divergent between 2010 and 2011. In 2011, PEBB/UMP transitioned to Regence as the insurance carrier, which may account for some differences. PEBB/UMP figures for 2010 also included a population covered by an alternative policy with Aetna, which may have had more health challenges and different reporting requirements. In order to correctly report current trends, average percent change is calculated using only for 2011-2013.

Few L&I claims during 2010-2013 included breast imaging tests:

Number of Claims	Number of Service Dates	Total Paid	Average Paid per Service Date
127	143	$21,191	$148

Footnotes:

*Average Percent Change was adjusted for changes in the PEBB/UMP population

** 4 year overall totals for members are unique members within the 2010-2013 timeframe, and therefore may not be the total of annual counts

1 Breast Cancer Diagnoses – Count of unique members per year with any diagnosis code 174.x or 175.x.

2 Inconclusive Mammogram – Count of unique members per year with any diagnosis code in:

Diagnosis Code	Diagnosis Code Description	% reported 2010-2013
793.8	AB MAMMOGRAM NOS	39.1%
793.81	MAMMOGRAPHIC MICROCALCIF	11.7%
793.82	INCONCLUSIVE MAMMOGRAM	7.0%
793.89	AB FINDINGS-BREAST NEC	42.1%

3 Non digital Mammograms – CPT codes 77055/77056 reported without any digital image codes (G0202/4/6), in combination with 77051 (diagnostic) or 77052 (screening), but about 60% without this specification.

4 Digital Mammograms – CPT codes 77055, 77056, 77051, 77052 with digital imaging codes G0202 (screening), G0204 or G0206 (diagnosis).

5 Tomographic Breast Imaging – CPT code 76499 (unlisted) with 76376/76377 interpretation, 77052 screening mammogram and a digital imaging code (G0202/4/6).

6 Ultrasound Breast Imaging - CPT codes 76645 - automated with 76376/76377 interpretation. Automated ultrasound procedures numbered 5, 6 and 10 in 2011, 2012, and 2013 respectively, so were not reported separately

7 Breast MRI - CPT codes 77058/77059 with 77051/77052 (screening or diagnostic) and 76376/76377 interpretation (optional).

8 Biopsy (non-imaging comparator) – CPT codes 19102, 19103 or 19125 with 19291, 19292, 19295 guidance, 88305/88307 pathology exams, 77021, 77031/2, 77042 guidance with digital imaging (G0204/G0206) or radiologic exam (76098)

Figure 1b. Medicaid Breast Cancer Imaging Counts and Costs, 2010-2013

Medicaid, Fee For Service Clients(FFS) & Managed Care Organizations (MCO)		2010	2011	2012	2013	4 Yr Overall **	Avg Annual % Change	
Populations	Average Annual Clients (FFS+MCO)	1,155,461	1,168,947	1,207,977	1,242,794		2.5%	
	Women (FFS+MCO)	841,640	856,812	853,551	864,935		-1.5%	*
	Women 40-74 (FFS+MCO)	132,328	138,321	142,392	148,721		1.5%	*
	Breast Cancer Diagnoses (FFS+MCO) [1]	2686	3171	3139	3271	6179	4.6%	*
	Inconclusive Mammogram(FFS+MCO) [2]	3104	3599	3371	3453	11,701	1.6%	*
Breast Imaging								
All Breast Imaging	**Total Paid (FFS)**	**$1,501,876**	**$1,517,347**	**$1,199,194**	**$667,285**	**$4,885,703**	-20.1%	*
	Procedures (FFS+MCO)	28,652	29,321	27,079	27,875	112,927	-3.1%	*
Non Digital Mammograms[3]	**Paid (FFS)**	**$458,407**	**$407,858**	**$326,847**	**$189,785**	**$1,382,897**	-22.9%	*
	Average Paid (FFS)	$76	$74	$84	$91	$81	6.6%	
	Procedures (FFS+MCO)	9,224	8,809	7,715	7,139	32,887	-10.3%	*
	% Screening CPT code (cts)	18.5%	14.2%	12.7%	5.2%			
	% Diagnostic CPT code (cts)	1.5%	1.6%	0.9%	0.3%			
Digital Mammograms[4]	**Paid (FFS)**	**$699,932**	**$756,869**	**$561,943**	**$271,165**	**$2,289,909**	**-22.0%**	*
	Average Paid (FFS)	$82	$90	$90	$99	$90	6.8%	
	Procedures (FFS+MCO)	12,007	12,381	11,432	11,896	47,716	-2.5%	*
	% Screening CPT code (cts)	98.4%	98.1%	98.8%	99.5%			
	% Diagnostic CPT code (cts)	1.6%	1.9%	1.2%	0.5%			
Tomographic Breast Imaging[5]	**Paid $ (FFS)**	**0**	**$5**	**$29**	**$860**	**$895**	1180.8%	*
	Average Paid (FFS)	0	$5	$29	$41	$19	174.3%	
	Procedures (FFS + MCO)	0	10	199	1034	1243	743.6%	*
US Breast Imaging[6]	**Paid $ US (FFS)**	**$250,504**	**$256,998**	**$210,588**	**$134,237**	**$852,326**	**-16.7%**	*
	Average Paid (FFS)	$65	$70	$70	$69	$69	-0.6%	
	Procedures (FFS + MCO)	7,134	7,765	7,333	7,403	29,635	-3.5%	*
Breast MRI[7]	**Paid $ BMRI (FFS)**	**$93,034**	**$95,617**	**$99,787**	**$71,238**	**$359,676**	-6.5%	*
	Average Paid (FFS)	$443	$376	$346	$311	$369	-6.1%	
	Procedures (FFS + MCO)	287	356	400	403	1,446	2.2%	*
Biopsy (Non-imaging comparator)[6]								
	Paid $ Bx (FFS)	**$509,665**	**$594,980**	**$532,757**	**$371,862**	**$2,009,264**	-12.0%	*
	Average Paid (FFS)	$441	$498	$509	$486	$484	-0.8%	
	Procedures (FFS + MCO)	1,666	1,991	1,875	1,848	7,380	-4.4%	*

Notes:

Payments for services are not reported for managed care organizations, paid amounts are not available for reporting. As client populations move to managed care, reported payments will drop, accounting for strong negative trends in Medicaid reported payments.

Client counts are not reported because they are almost the same as the procedure counts. Annual usage rates for most procedures are less than 1.1 procedures per member per year.

Footnotes:

*Average Percent Change was adjusted for changes in the Medicaid population

** 4 year overall totals for members are unique members within the 2010-2013 timeframe, and therefore may not be the total of annual counts

1 Breast Cancer Diagnoses – Count of unique members per year with any diagnosis code 174.x or 175.x.

2 Inconclusive Mammogram – Count of unique members per year with any diagnosis code in:

Diagnosis Code	Diagnosis Code Description	% reported 2010-2013
793.8	AB MAMMOGRAM NOS	34.4%
793.81	MAMMOGRAPHIC MICROCALCIF	16.0%
793.82	INCONCLUSIVE MAMMOGRAM	4.6%
793.89	AB FINDINGS-BREAST NEC	45.0%

3 Non digital Mammograms – CPT codes 77055/77056 reported without any digital image codes (G0202/4/6), in combination with 77051 (diagnostic) or 77052 (screening), but about 40% without this specification.

4 Ultrasound Breast Imaging - CPT codes 76645 - automated with 76376/76377 interpretation. Automated ultrasound procedures numbered 7 in 4 years, so were not reported separately

5 Breast MRI - CPT codes 77058/77059 with 77051/77052 (screening or diagnostic) and 76376/76377 interpretation (optional).

6 Biopsy (non-imaging comparator) – CPT codes 19102, 19103 or 19125 with 19291, 19292, 19295 guidance, 88305/88307 pathology exams, 77021, 77031/2, 77042 guidance with digital imaging (G0204/G0206) or radiologic exam (76098)

Figure 2: Agency Fee Schedules

Current pricing for Breast Imaging components as available on agency web sites:

CPT Codes	CPT Code Descriptions	Current Agency Fees (Allowed)		
		PEBB/UMP*	L&I†	Medicaid‡
76376	3D rendering with interpretation (no independent workstation)	$94.57	$49.21	$17.67
76377	Requiring image post processing on independent workstation	$124.00	$141.76	$50.90
76499	Unlisted diagnostic radiographic procedure	Not listed	By report	By report
76645	Ultrasound, breast(s)	$140.89	$169.88	$61.00
77051	Computer-aided detection; diagnostic	$15.44	Not covered	$6.31
77052	screening mammography	$15.44	Not covered	$6.31
77055	Mammography; unilateral	$124.49	$152.89	$54.90
77056	Bilateral	$159.23	$196.24	$70.46
77057	Screening mammography, bilateral	$115.32	$140.01	$50.27
77058	Magnetic resonance imaging, breast, unilateral	$997.33	$948.41	$340.52
77059	Bilateral	$1,016.14	$944.90	$339.26
G0202	Screening mammography, produces direct digital image, bilateral, all views	198.31	$230.22	$82.66
G0204	Diagnostic mammography, producing direct digital image, bilateral, all views	239.32	$280.60	$100.75
G0206	Diagnostic mammography, producing direct digital image, unilateral, all views	189.14	$220.85	$79.29

*Regence Blue Shield Provider Fee Schedule – effective January 1 2013, MD/DO/DPM Provider rates, Maximum Allowable fee, http://www.hca.wa.gov/ump/documents/Regence_Professional_Fee_Schedule_Jan_2013.pdf, Accessed 11/3/2014. Payment based on the Regence Fee Schedule is subject to all of the terms and conditions of the applicable Regence BlueShield provider agreement, member benefits, Regence BlueShield policies, and all published Regence BlueShield administrative guidelines. Therefore, the appearance of fees for particular procedure codes does not guarantee coverage. Some providers may have contracted fees at different rates.

†Washington State Labor and Industries Fee Schedules and Payment Policies (MARFS), Fee Schedules and Payment Policies for: 2014, http://www.lni.wa.gov/apps/FeeSchedules/, Accessed 11/3/2014

‡Washington State Medicaid Rates Development Fee Schedule, July 1, 2014 Physician and Related Services Fee Schedule (Updated October 1, 2014), http://www.hca.wa.gov/medicaid/rbrvs/pages/index.aspx#P, Accessed 11/3/2014.

Figure 3a. PEBB/UMP Inconclusive Mammogram Diagnoses by Age, 2010-2013

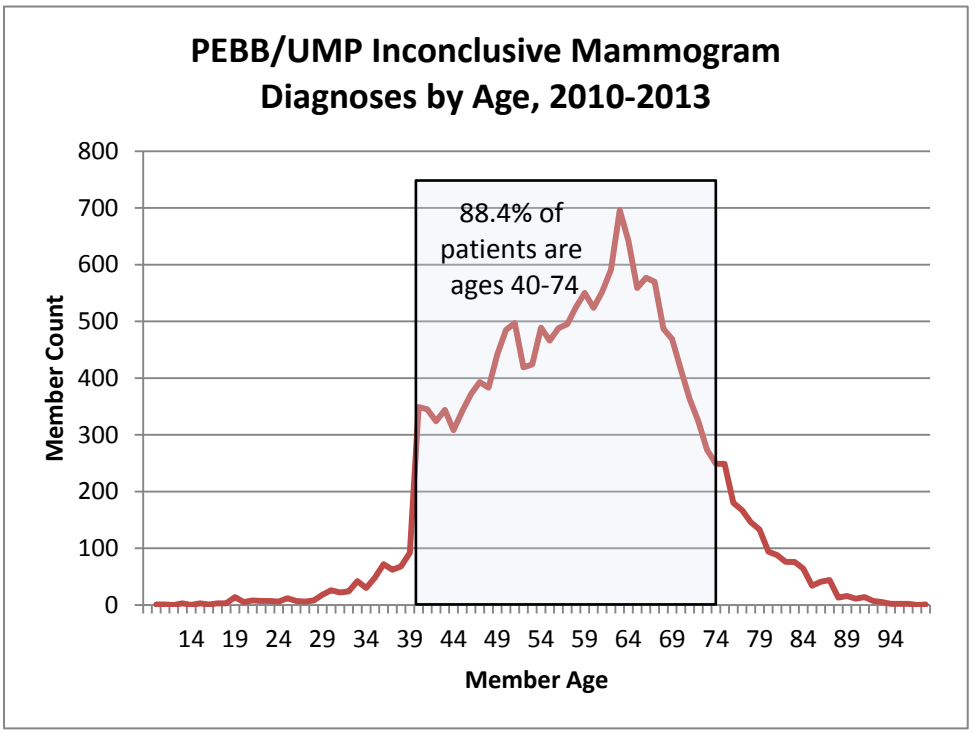


Figure 3b. Medicaid Inconclusive Mammogram Diagnoses by Age, 2010-2013

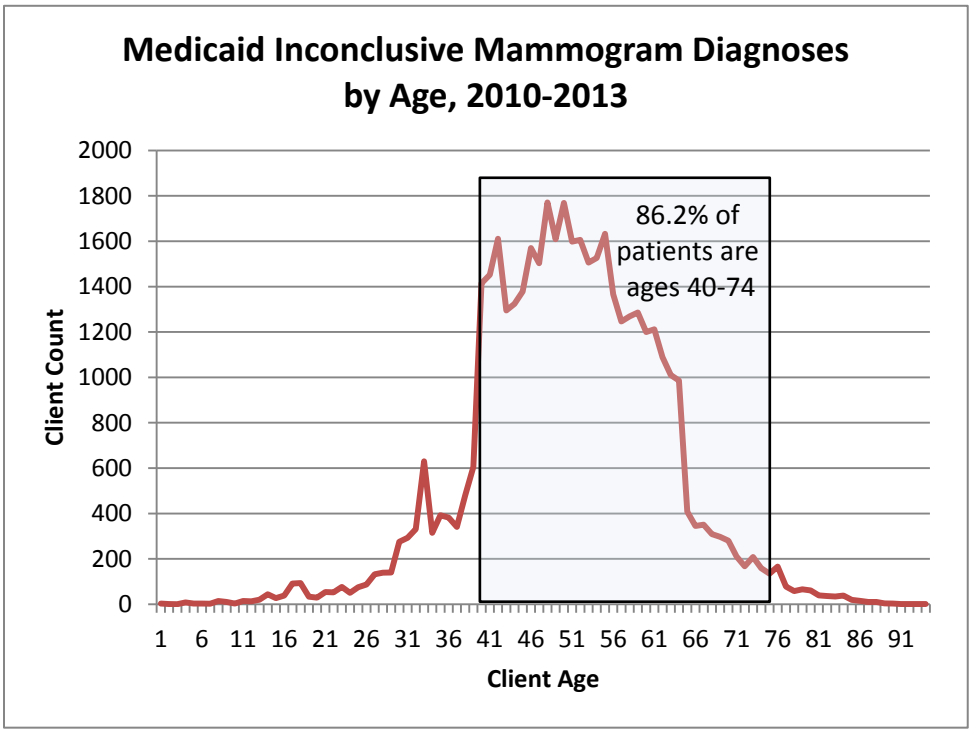
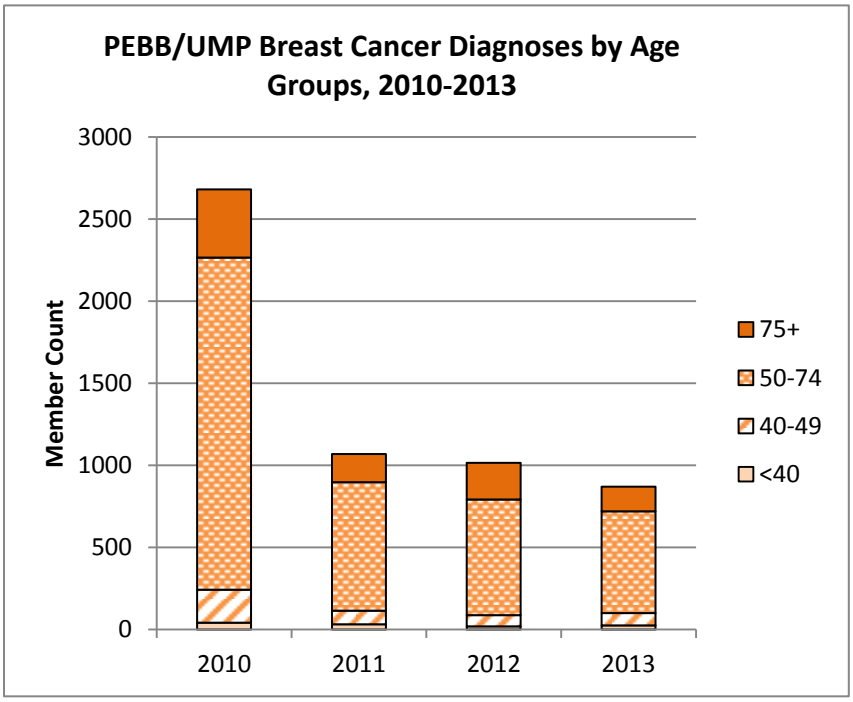


Figure 4. Breast Cancer Diagnoses by Age Group

Figure 4a. PEBB UMP Breast Cancer Diagnoses by Age Group, 2010-2013



NOTE: Some PEBB/UMP statistics appear to be widely divergent between 2010 and 2011. In 2011, PEBB/UMP transitioned to Regence as the insurance carrier, which may account for some differences. PEBB/UMP figures for 2010 also included a population covered by an alternative policy with Aetna, which may have had more health challenges and different reporting requirements.

Figure 4b. Medicaid Breast Cancer Diagnoses by Age Group, 2010-2013

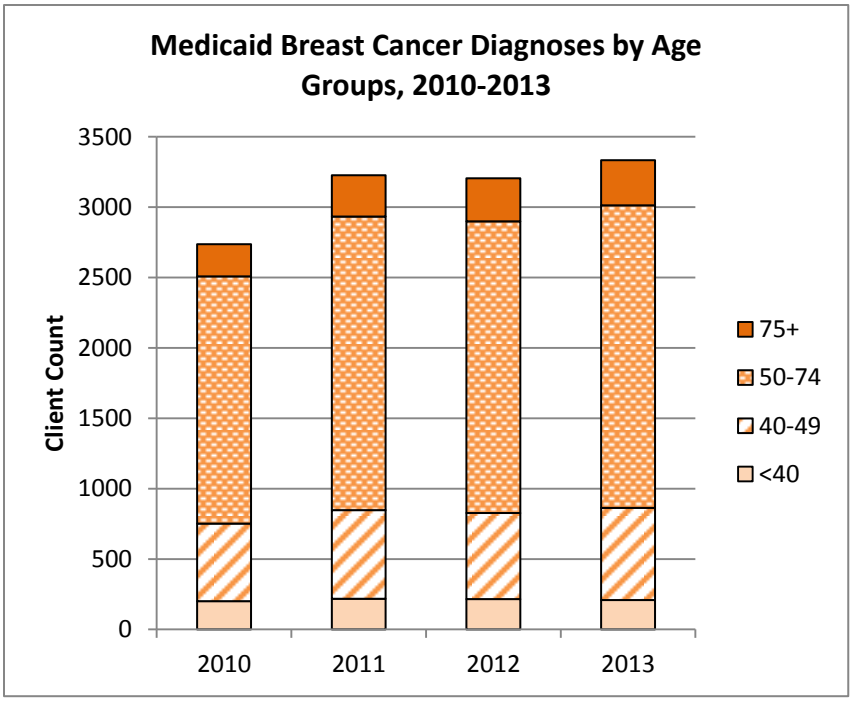


Figure 5. Services by Age Group

Figure 5a. PEBB/UMP Mammograms by Age Groups, 2010-2013

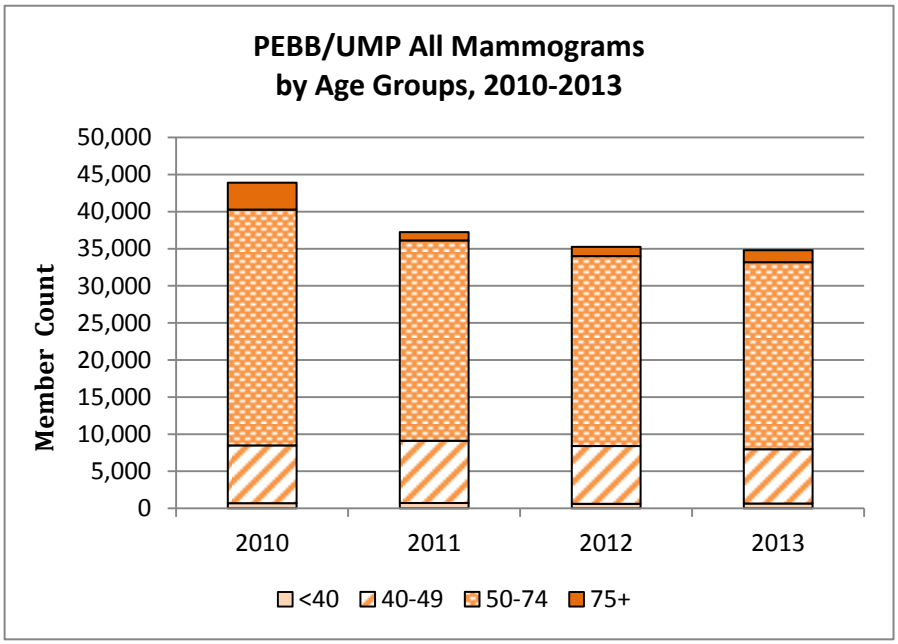


Figure 5b. PEBB/UMP Non-digital Mammograms by Age Groups 2010-2013

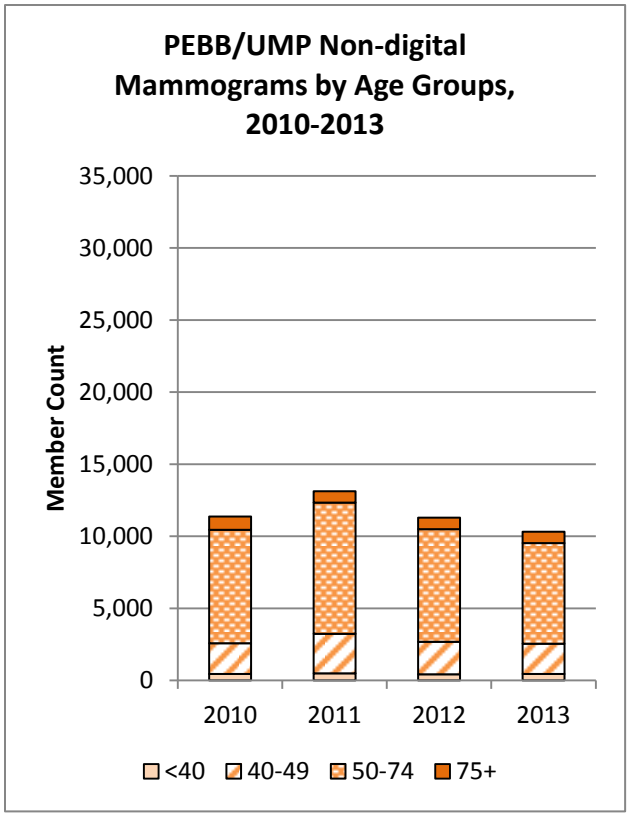


Figure 5c. PEBB/UMP Digital Mammograms by Age Groups, 2010-2013

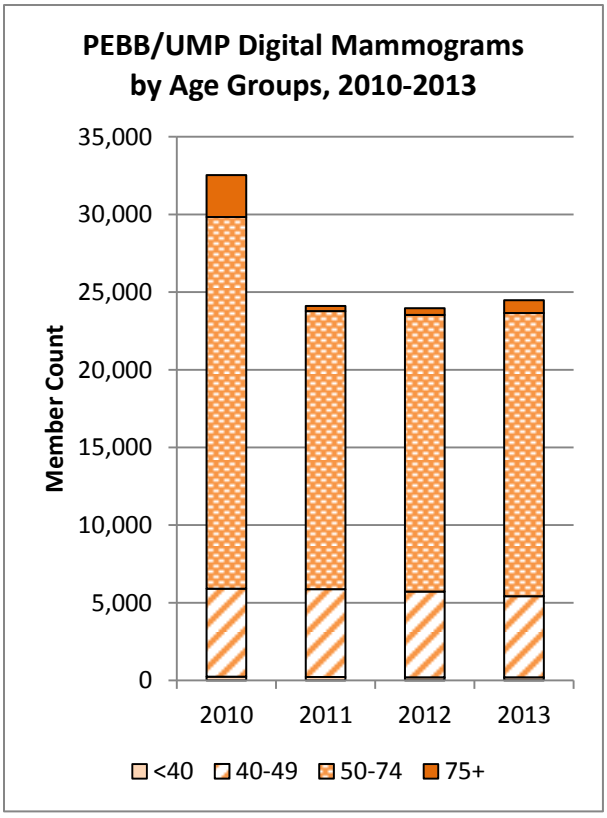


Figure 5d. PEBB/UMP Tomography by Age Groups, 2010-2013

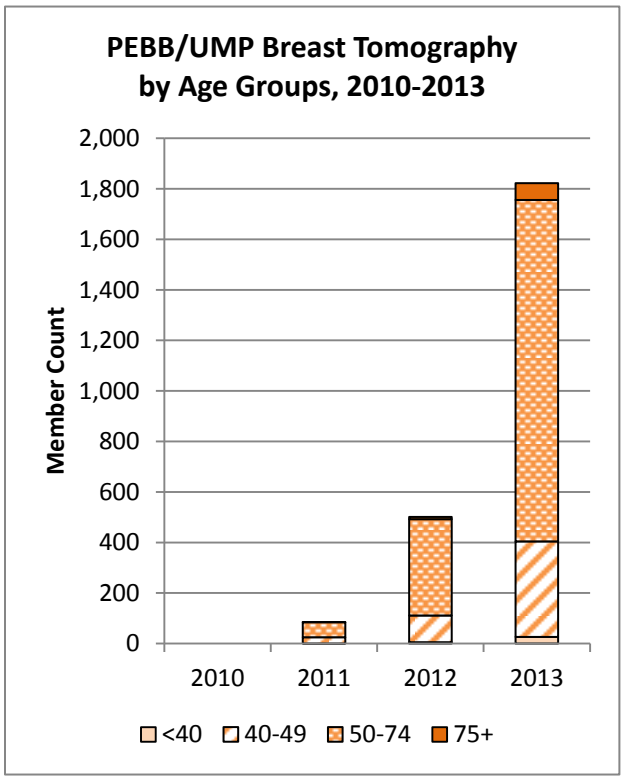


Figure 5e. PEBB/UMP Breast MRI by Age Groups, 2010-2013

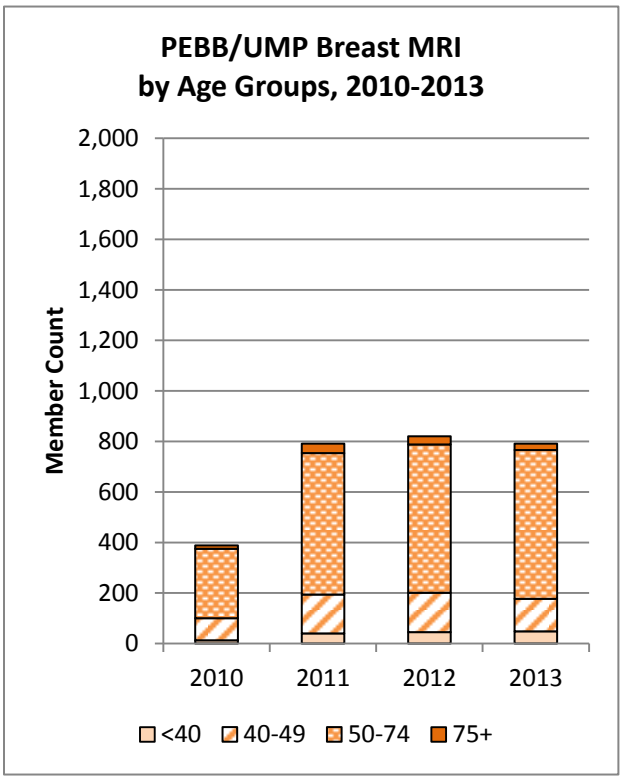


Figure 5f. PEBB/UMP Breast Ultrasound by Age Groups, 2010-2013

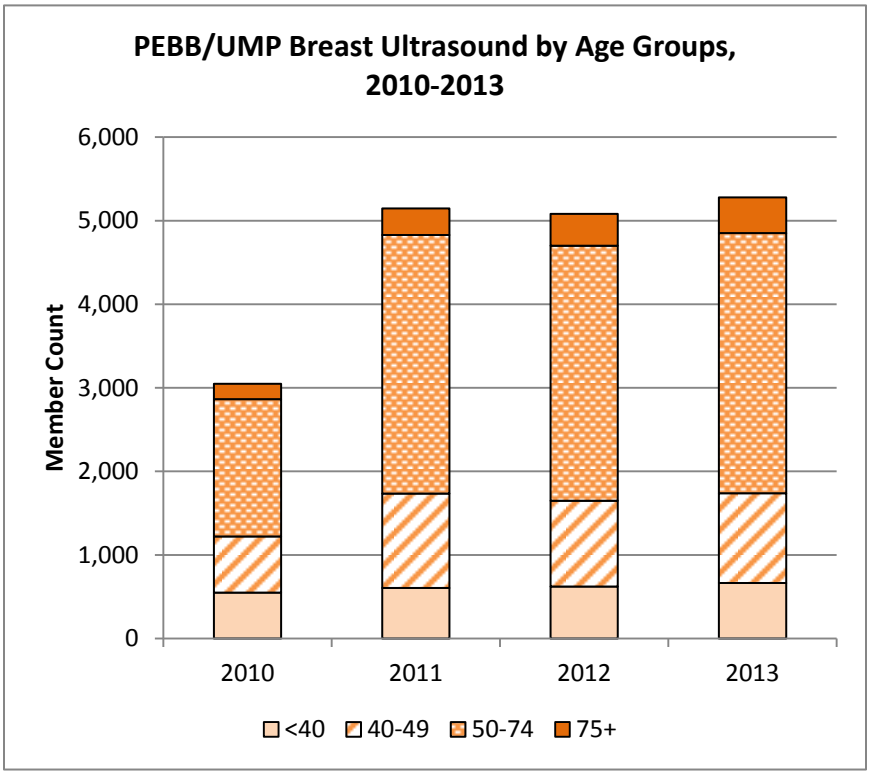


Figure 5g. Medicaid Mammograms by Age Groups, 2010-2013

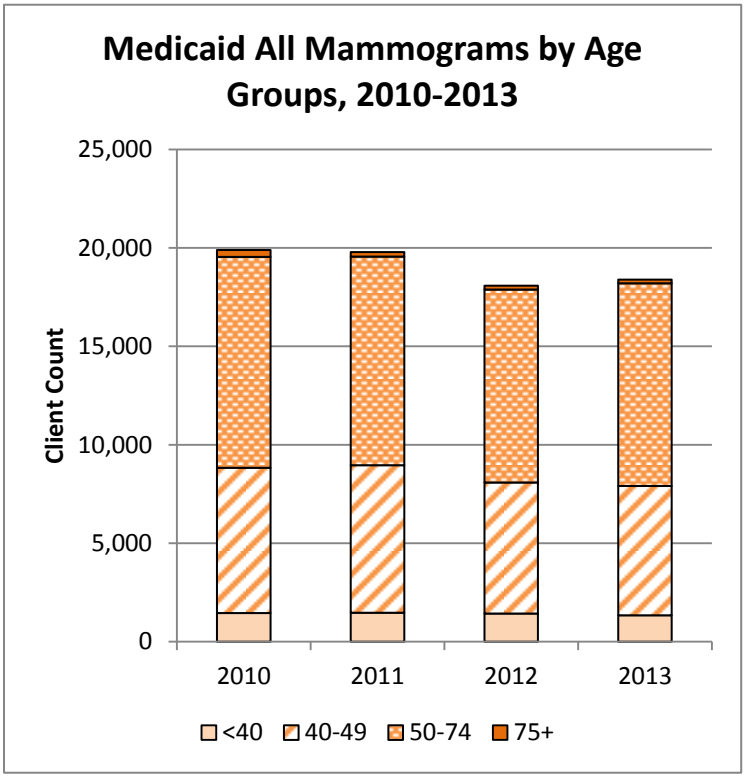


Figure 5h. Medicaid Non-Digital Mammograms by Age Groups

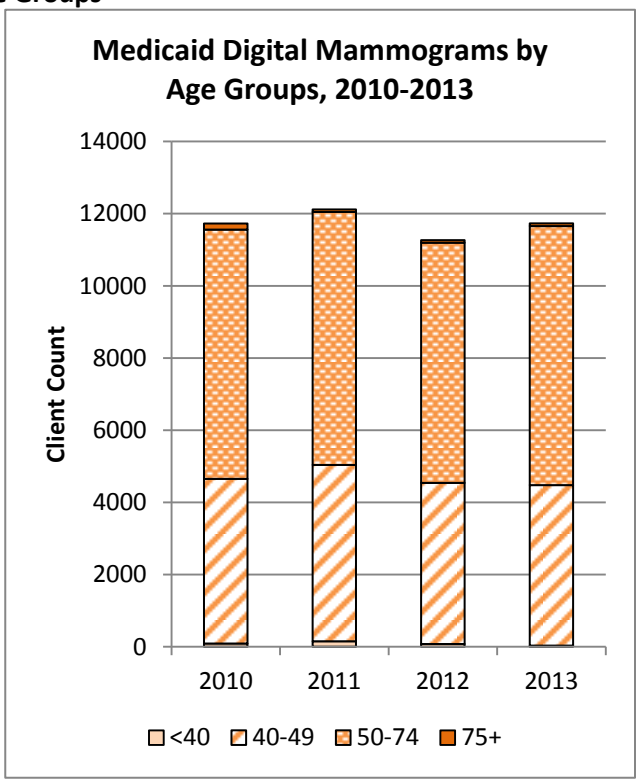


Figure 5i. Medicaid Digital mammograms by Age Groups

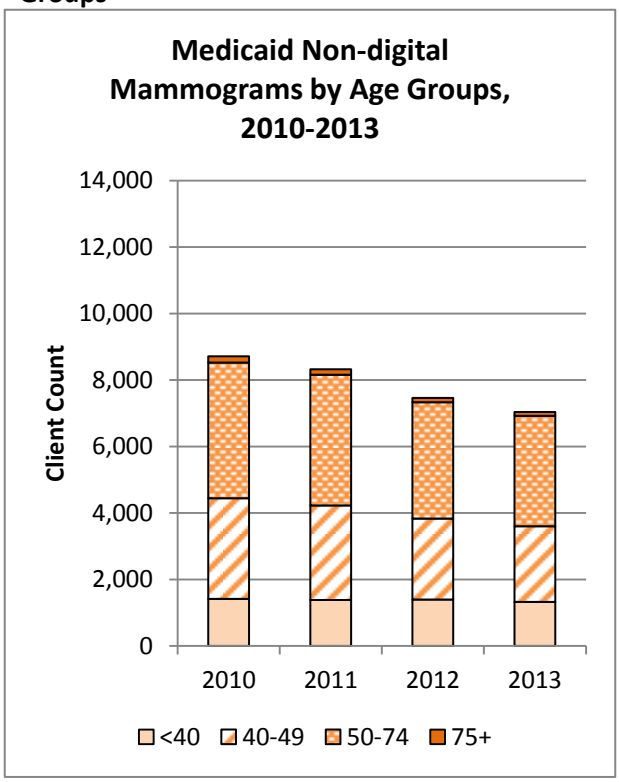


Figure 5j. Medicaid Tomography by Age Groups, 2010-2013

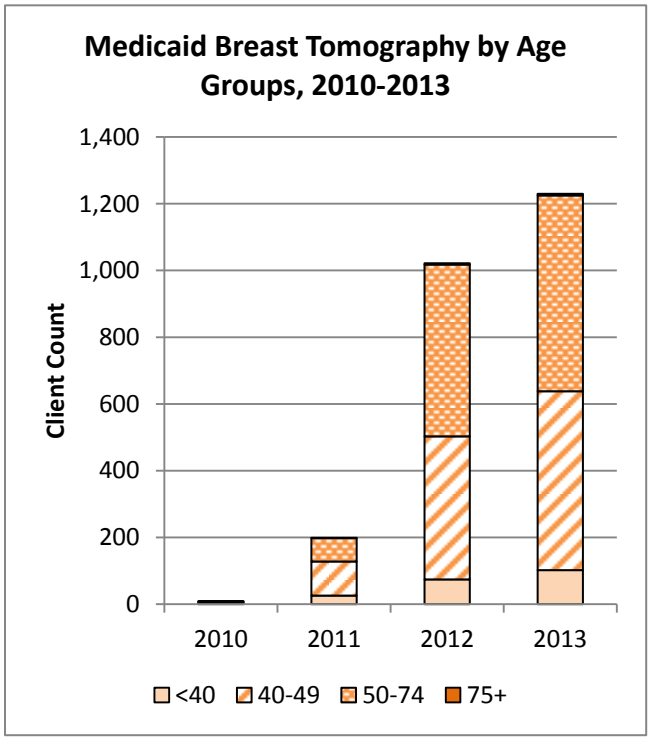


Figure5k. Medicaid BMRI by Age Groups, 2010-2013

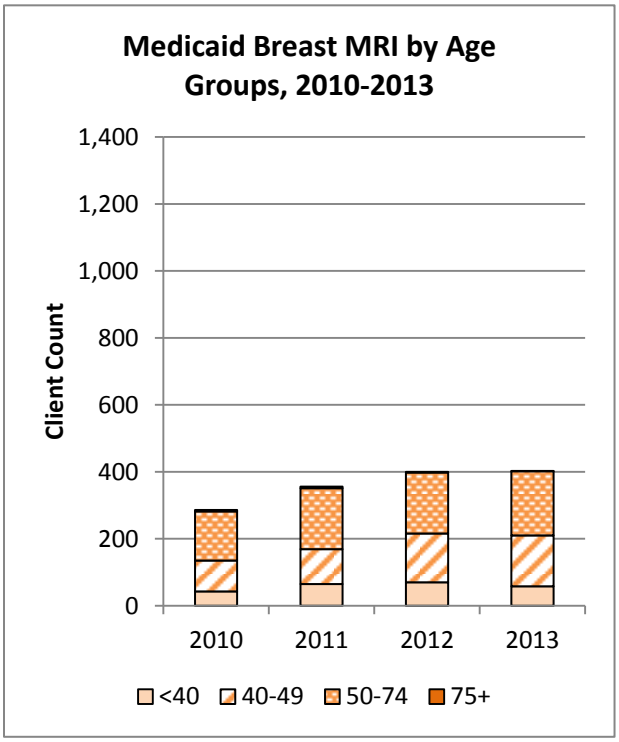


Figure 5l. Medicaid Breast Ultrasound by Age Groups, 2010-2013

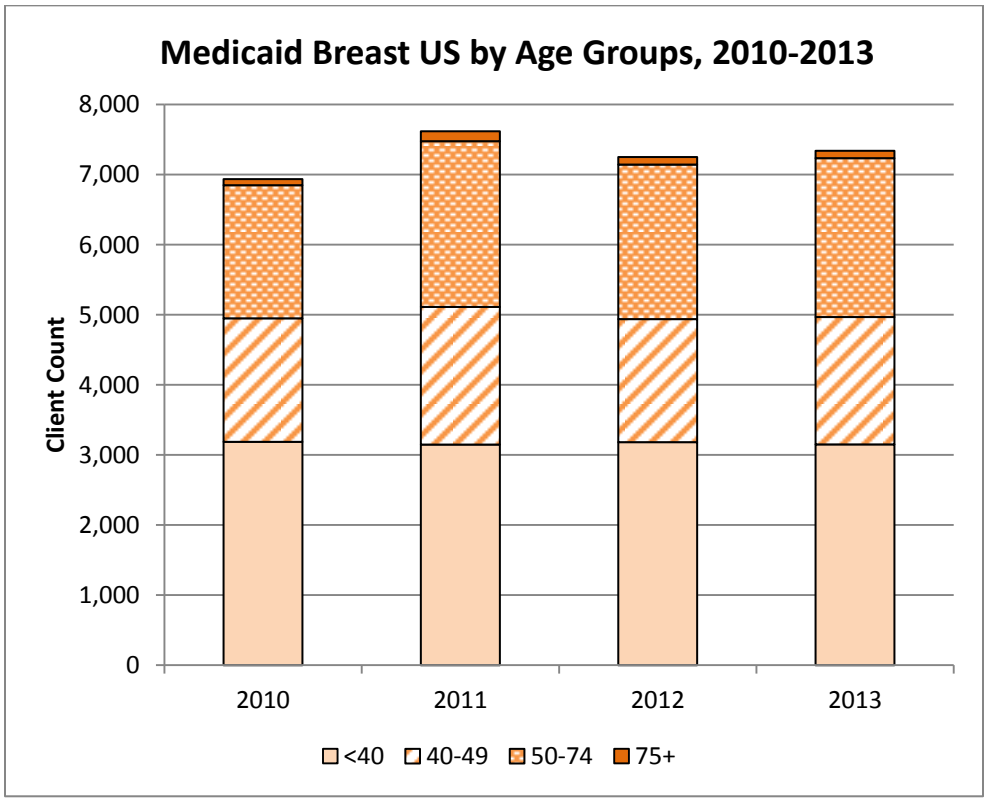


Figure 6 Screening and Diagnostic Services in Patients with Inconclusive Finding by Age Group, 2010-2013

Figure 6a. PEBB/UMP Screening/ Diagnostic Procedure Counts for Members with Inconclusive Findings, <40, 2010-2013

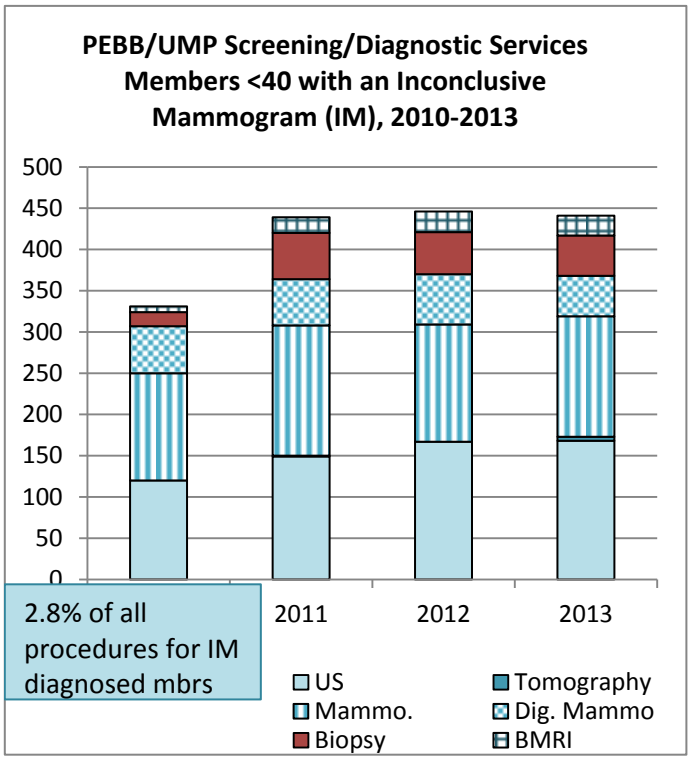


Figure 6b. PEBB/UMP Screening/ Diagnostic Procedure Counts for Members with Inconclusive Findings, 40-49, 2010-2013

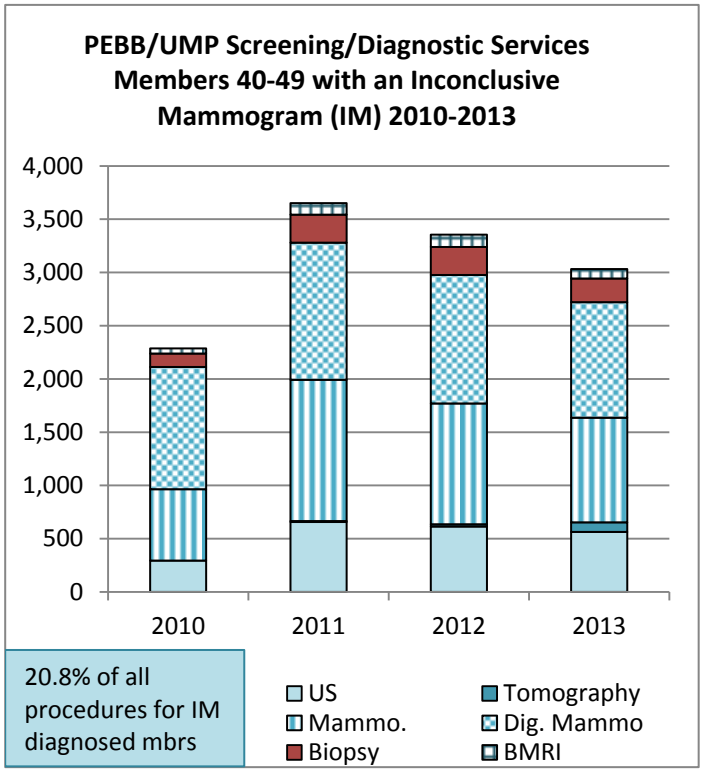


Figure 6c . PEBB/UMP Screening/Diagnostic Procedure Count for Members with Inconclusive Findings, 50-74, 2010-2013

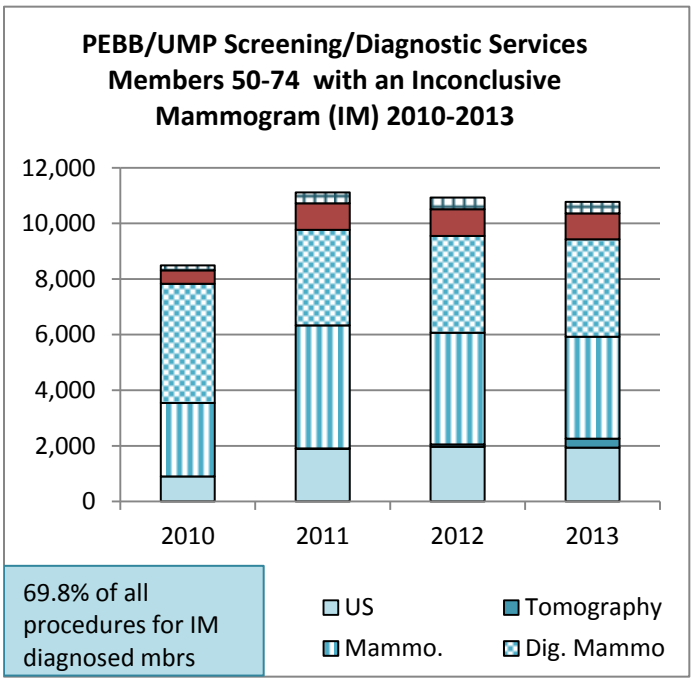


Figure 6d . PEBB/UMP Screening/Diagnostic Procedure Counts for Members with Inconclusive Findings, 75+, 2010-2013

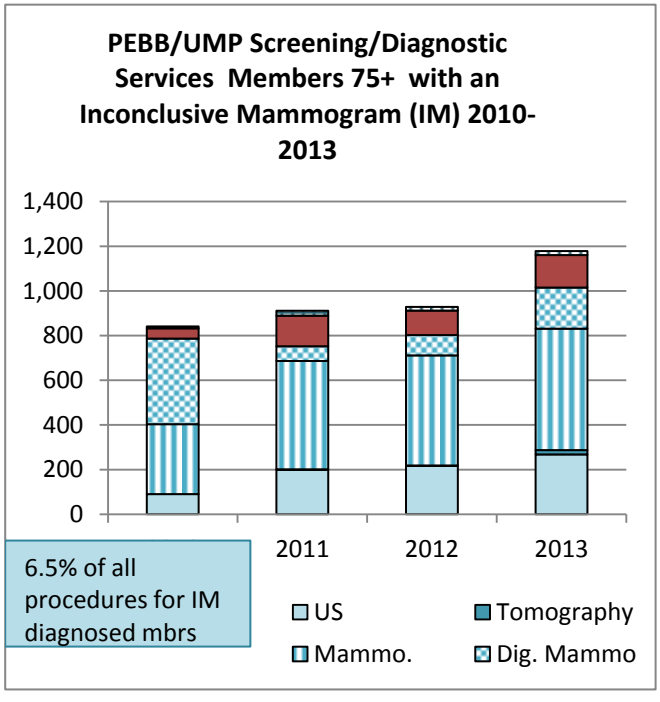


Figure 6e. Medicaid Screening and Diagnostic Procedures in Clients with Inconclusive Findings, <40, 2010-2013

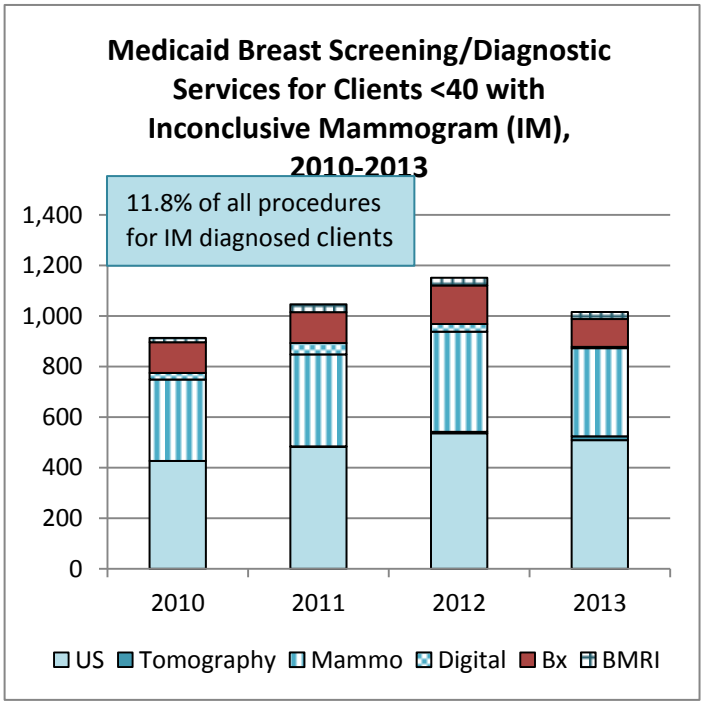


Figure 6f . Medicaid Screening and Diagnostic Procedures in Clients with Inconclusive Findings, 40-49, 2010-2013

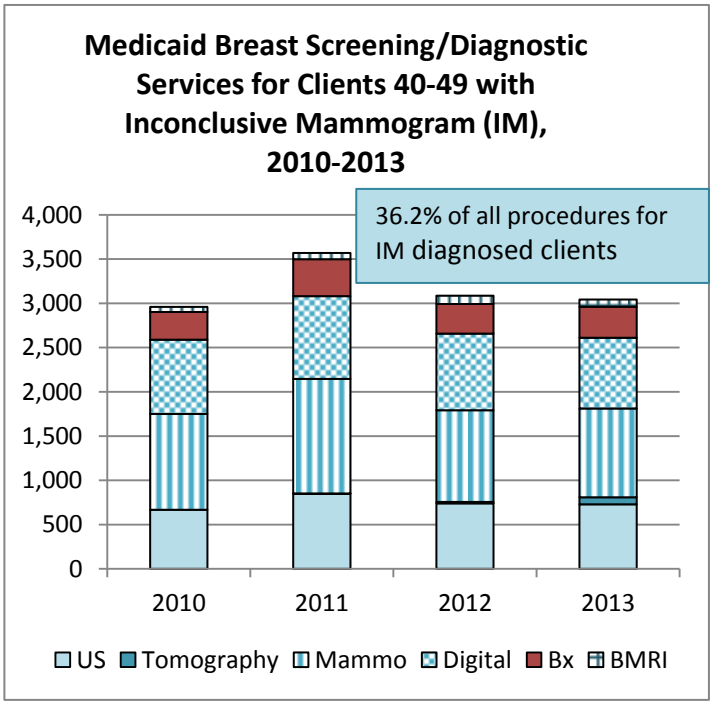


Figure 6g . Medicaid Screening and Diagnostic Procedures in Clients with Inconclusive Findings, 50-74, 2010-2013

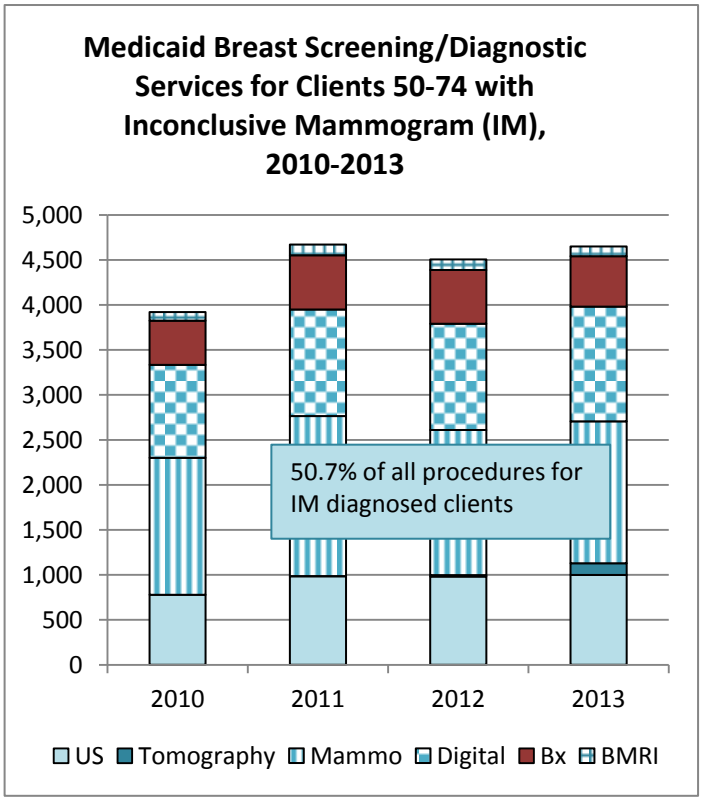
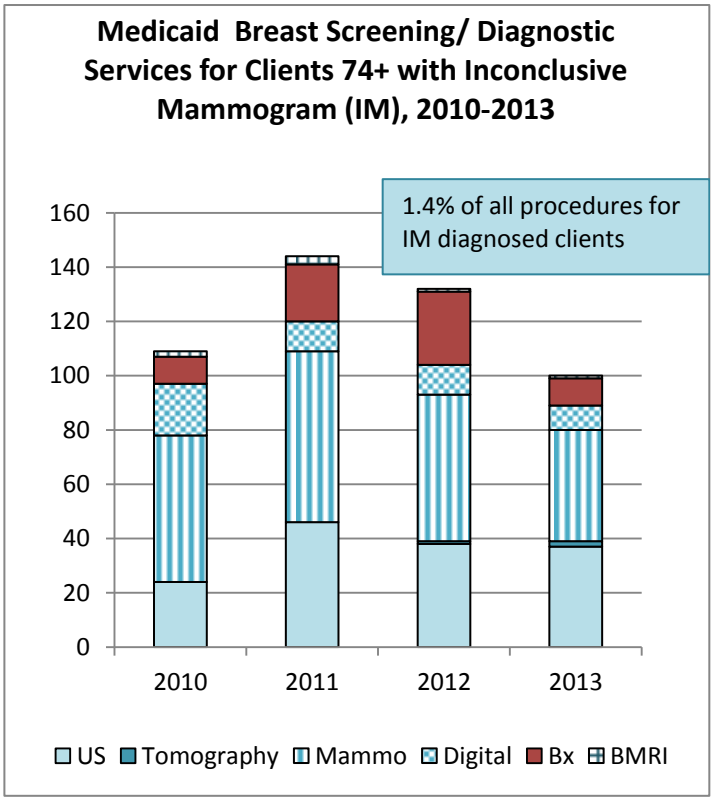


Figure 6h. Medicaid Screening and Diagnostic Procedures in Clients with Inconclusive Findings, 74+, 2010-2013



CPTs Code	Related Medical Codes
76376	3D rendering with interpretation and reporting of computed tomography, magnetic resonance imaging, ultrasound, or other tomographic modality with image postprocessing under concurrent supervision; not requiring image postprocessing on an independent workstation (Use 76376 in conjunction with code[s] for base imaging procedure[s]) (Do not report 76376 in conjunction with 31627, 70496, 70498, 70544-70549, 71275, 71555, 72159, 72191, 72198, 73206, 73225, 73706, 73725, 74174, 74175, 74185, 74261-74263, 75557, 75559, 75561, 75563, 75565, 75571-75574, 75635, 76377, 78012-78999, 0159T)
76377	requiring image postprocessing on an independent workstation (Use 76377 in conjunction with code[s] for base imaging procedure[s]) (Do not report 76377 in conjunction with 70496, 70498, 70544-70549, 71275, 71555, 72159, 72191, 72198, 73206, 73225, 73706, 73725, 74174, 74175, 74185, 74261-74263, 75557, 75559, 75561, 75563, 75565, 75571-75574, 75635, 76376, 78012-78999, 0159T)
76499	Unlisted diagnostic radiographic procedure (For electrical impedance breast scan, use 76499)
76645	Ultrasound, breast(s) (unilateral or bilateral), real time with image documentation
77051	Computer-aided detection (computer algorithm analysis of digital image data for lesion detection) with further review for interpretation, with or without digitization of film radiographic images; diagnostic mammography (List separately in addition to code for primary procedure) (Use 77051 in conjunction with 77055, 77056)
77052	screening mammography (List separately in addition to code for primary procedure) (Use 77052 in conjunction with 77057)
77055	Mammography; unilateral
77056	Bilateral (Use 77055, 77056 in conjunction with 77051 for computer-aided detection applied to a diagnostic mammogram)
77057	Screening mammography, bilateral (2-view film study of each breast) (Use 77057 in conjunction with 77052 for computer-aided detection applied to a screening mammogram)
77058	Magnetic resonance imaging, breast, without and/or with contrast material(s); unilateral
77059	bilateral

CPTs Code	Description and Coding Instructions
G0202	Screening mammography, produces direct digital img, bilaterl, all views
G0204	Diagnostic mammography, producing direct digital image, bilateral, all views
G0206	Diagnostic mammography, producing direct digital image, unilateral, all views
19102191033 19125	Percutaneous wo/ with automated vacuum assted or rotating biopsy dev, using imaging guidance, Deleted for 2014
192901929 1	Preoperative placement of needle localization wire, breast, first lesions, each additional lesion
19295	Image guided placement, metallic localization clip, percutaneous, during breast biopsy/aspiration
76098	Radiological examination surgical specimen
76942	US guidance for needle placement (eg biopsy, aspiriation, injection, localization device), imaging supervision and interpretation
77021	MRI guidance for needle placement (eg biopsy, aspiriation, injection, localization device), imaging supervision and interpretation
7703177032	Stereotactic (77031) or mammographic (77032) guidance for needle placement (eg biopsy, aspiriation, injection, localization device), imaging supervision and interpretation, Deleted for 2014
88305 88307	Surgical pathology, gross and microscopic exam, level 4 (88305) and level 5 (88307)
ICD9	**Related Diagnoses**
793.8x	Abnormal non-specific finding: Breast
793.80	Abnormal mammogram, unspecified
793.81	Mammographic microcalcification, excluding mammographic calcification (793.89) and mammographic calculus (793.89)
793.82	Inconclusive mammogram: Dense breasts NOS Inconclusive mammogram NEC Inconclusive mammography due to dense breasts Inconclusive mammography NEC
793.89	Other abnormal findings on radiological examination of breast Mammographic calcification Mammographic calculus
174.xx	All female breast cancer diagnoses
175.xx	All male breast cancer diagnoses

2. Clinical Guidelines and Training Standards

2.1 Magnetic Resonance Imaging (MRI) of the Breast

The American Cancer Society (ACS) (2014)

http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs

The ACS recommends annual adjunctive MRI for women at high risk for breast cancer. This includes women whose lifetime risk of breast cancer is 20% to 25% or greater; women who have a known *BRCA1* or *BRCA2* gene mutation, or who have a first-degree relative with these genetic mutations if they have not been tested themselves; women who had radiation therapy to the chest between ages 10 and 30; and women who have Li-Fraumeni syndrome, Cowden syndrome, or Bannayan-Rilyey-Ruvalcaba syndrome. The ACS recommends against MRI screening for women with a low lifetime risk of breast cancer, defined as less than 15%. The society suggests that there is not enough evidence to form MRI recommendations for women with moderate risk of developing breast cancer, or who may be at increased risk for breast cancer due to factors such as having extremely or heterogeneously dense breast tissue on mammogram, a personal history of breast cancer, ductal carcinoma in situ, lobular carcinoma in situ, atypical ductal hyperplasia, or atypical lobular hyperplasia.

National Comprehensive Cancer Network (NCCN) (2014)

http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf

The NCCN recommends MRI as an adjunct to mammography starting at age 30 for women with a lifetime risk of breast cancer greater than 20% (using Claus, BRCAPRO, BOADICEA, or Tyrer-Cuzick models), as well for women with mutations in BRCA1, BRCA2, TP53, or PTEN and their untested first-degree relatives. In addition, they recommend annual screening MRI for those receiving radiation therapy to their chest between the ages of 10 to 30 years starting 8 to 10 years following the radiation therapy or at age 40, whichever comes first.

The NCCN guidelines also state that there is insufficient evidence to recommend for or against annual MRI screening for the following women: those with a 15% to 20% lifetime risk for breast cancer; those with a personal history of breast cancer, ductal carcinoma in situ, lobular carcinoma in situ, atypical ductal hyperplasia, atypical lobular hyperplasia; or those with heterogeneously dense or extremely dense tissue on mammography. NCCN recommends against MRI for women with a lifetime risk of less than 15%.

American College of Radiology / Society of Breast Imaging (2010)

http://www.jacr.org/article/S1546-1440(09)00480-3/fulltext

Joint guidelines from the American College of Radiology and the Society of Breast imaging recommend annual screening MRI examinations starting at age 30 for BRCA mutation carriers and their untested first degree relatives, for women with greater than a 20% lifetime risk for breast cancer on the basis of family history, women with a history of chest irradiation (usually for Hodgkin's disease), and a single screen of the contralateral breast for women with newly diagnosed breast cancer (Lee et al., 2010). They recommend considering screening MRI for women with a lifetime risk between 15% and 20% on the

basis of a personal history of breast or ovarian cancer or biopsy proven lobular neoplasia or atypical ductal hyperplasia.

The European Society of Breast Imaging (2007)

http://www.eusobi.org/html/img/pool/330_2008_863_OnlinePDF.PDF

The European Society of Breast Imaging recommends annual MRI screening examinations for women with a BRCA mutation, first degree relatives of BRCA carriers, women with radiation to their chest wall between the ages of 10 and 30 years, women with Li-Fraumeni syndrome (TP53 mutation carriers) and their untested first degree relatives, and women with Cowden syndrome (PTEN mutation carriers) and their first degree relatives) (Mann et al., 2007).

2.2 Hand-held Ultrasonography (HHUS) of the Breast

The American Cancer Society (ACS) (2014)

http://www.cancer.org/cancer/breastcancer/moreinformation/breastcancerearlydetection/breast-cancer-early-detection-acs-recs

The ACS has no recommendation on HHUS for breast cancer screening.

National Comprehensive Cancer Network (NCCN) (2014)

http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf

Under breast screening considerations, the NCCN guidelines state “Dense breasts limit the sensitivity of mammography. Dense breasts are associated with an increased risk for breast cancer, but there is insufficient evidence to support routine supplemental screening in women with dense breasts and no other risk factors” (NCCN, 2013). Under the same section they also note, “There are several studies supporting the use of ultrasound for breast cancer screening as an adjunct to mammography for high risk women with dense breast tissue.”

American College of Radiology / Society of Breast Imaging (2010)

http://www.jacr.org/article/S1546-1440(09)00480-3/fulltext

Joint guidelines from the American College of Radiology and the Society of Breast imaging recommend considering annual screening ultrasound examinations in addition to mammography for women eligible for MRI screening who cannot have MRI for any reason (Lee et al., 2010). They recommend considering ultrasound in women with dense breast tissue as an adjunct to mammography.

2.3 Automated Whole Breast Ultrasonography (ABUS)

There are no guidelines currently recommending ABUS to screen for breast cancer from any major clinical society, including the American Cancer Society, the National Comprehensive Cancer Network, the American College of Radiology, and the Society of Breast Imaging.

2.4 Digital Breast Tomosynthesis (DBT)

American Cancer Society (ACS) (2014)

http://www.cancer.org/healthy/findcancerearly/examandtestdescriptions/mammogramsandotherbreastimagingprocedures/mammograms-and-other-breast-imaging-procedures-improving-mammo

The ACS suggests that DBT **"**uses more radiation than most standard 2-view mammograms, but it may allow doctors to see [dense areas] more clearly. Some studies have suggested it might lower the chance that the patient will be called back for unnecessary tests. It may also be able to find more cancers." ACS does not provide a recommendation for or against use of DBT.

American College of Radiology (ACR) (2014)

http://www.acr.org/About-Us/Media-Center/Position-Statements/Position-Statements-Folder/20141124-ACR-Statement-on-Breast-Tomosynthesis

While digital mammography is the only breast cancer screening procedure that has been proven to reduce mortality, tomosynthesis is a very promising technology that has been shown to reduce recall rates and increase cancer detection, thus having a positive impact on patient care. The ACR acknowledges the lack of studies demonstrating long-term benefits, and encourages payers to reimburse for tomosynthesis so that additional large-scale studies can be conducted.

American Society of Breast Disease (ASBD) (2013)

https://www.asbd.org/news/ASBD_statement_on_Tomosynthesis12-16-13.pdf

Despite the growing use of full-field digital mammography over film, screening mammography is still limited by overlapping breast tissue. The use of DBT has the potential to overcome these limitations and improve diagnostic accuracy of breast cancer. DBT has the potential to improve patient outcomes, particularly with regards to diagnostic work-up following screening. By increasing cancer detection and reducing recalls, DBT has utility as both a diagnostic and screening tool and may have the greatest impact on women with dense breast tissue.

National Comprehensive Cancer Network (NCCN) (2014)

http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf

Under breast screening considerations, the NCCN guidelines state that "Early studies show promise for DBT mammography. Two large trials showing a combined use of digital mammography and tomosynthesis resulted in improved cancer detection and decreased call back rates; of note, this is double the dose of radiation and is a factor in recommending this modality. Definitive studies are still pending" (NCCN, 2013).

Washington State Radiological Society (WSRS) (2014)

http://www.wsrs.org/position_statements.html

Adding DBT to standard mammography screening programs would help overcome many of the current limitations of digital mammography, such as its inability to distinguish overlapping breast tissue. As evidenced by recent studies, the addition of DBT will likely result in the additional detection of one cancer for every 1,000 women screened. Moreover, it will reduce the number of unnecessary call-backs as well as decrease patient anxiety and lost productivity as a result of false-positive findings. DBT has the potential to both improve patient outcomes and decrease healthcare costs by identifying more

early-stage cancers and expediting diagnostic workup. WSRS urges payers to reimburse for DBT so this advancement in breast cancer screening can be more widely utilized.

3. Medicare and Representative Private Insurer Coverage Policies

3.1 Breast Ultrasound

Centers for Medicare and Medicaid Services (CMS)

The national coverage determination (NCD) for breast ultrasound relates only to its use for diagnosis rather than screening. There is also no current local coverage determination (LCD) for screening that covers the state of Washington. LCDs for Illinois (L26890) and Kentucky (L31856) on breast imaging relate only to breast ultrasound's diagnostic use.

Representative Regional Private Insurer Policies

Health Net / The Regence Group / Premera Blue Cross

No information regarding coverage for breast ultrasound screening was available from Health Net, Premera Blue Cross, or The Regence Group.

Representative National Private Insurer Policies

http://www.medsolutions.com/cignaguidelines/guideline_downloads/Chest%20Imaging%20Guidelines-Cigna.pdf

http://apps.humana.com/tad/Tad_New/Search.aspx?criteria=breast+imaging&searchtype=freetext&policyType=both

https://www.unitedhealthcareonline.com/ccmcontent/ProviderII/UHC/en-US/Assets/ProviderStaticFiles/ProviderStaticFilesPdf/Tools%20and%20Resources/Policies%20and%20Protocols/Medical%20Policies/Medical%20Policies/Breast_Imaging_for_Screening_and_Diagnosing_Cancer.pdf

CIGNA contracts MedSolutions, Inc. to administer its radiology benefits in Washington, and does not support the use of breast ultrasound as a screening tool. Humana and UnitedHealthcare consider breast ultrasound experimental and investigational for any type of breast cancer screening. No information regarding coverage for either HHUS or ABUS was publicly available from other national payers such as Aetna, UniCare, or Anthem/Wellpoint.

3.2 Breast MRI

Centers for Medicare and Medicaid Services (CMS)

The national coverage determination (NCD) for MRI relates only to its general diagnostic use rather than as a breast cancer screening method. There is no local coverage determination (LCD) for breast MRI screening that covers the state of Washington. LCDs for Illinois (L26890) and Kentucky (L31856) cover the use of breast MRI, but, as with ultrasound, indications are limited to diagnostic purposes only.

Representative Regional Private Insurer Policies

Health Net

https://www.healthnet.com/portal/provider/content/iwc/provider/unprotected/working_with_HN/content/medical_policies.action

Health net covers MRI as an adjunct to mammography for patients with familial history of breast cancer, genetic abnormalities, or in women who received radiation therapy to the chest between the ages of 10 and 30.

The Regence Group / Premera Blue Cross

http://aimspecialtyhealth.com/clinical-guidelines/guideline-set

Premera Blue Cross and The Regence Group contract AIM Specialty Health to conduct medical necessity reviews for radiologic procedures, including breast imaging. AIM Specialty Health's guidelines recommend MRI screening for patients with familial history of breast cancer, genetic abnormalities, or in women who received radiation therapy to the chest between the ages of 10 and 30. AIM Specialty Health also considers MRI screening medically necessary for women with lobular carcinoma in situ (LCIS) or ductal carcinoma in situ (DCIS).

Representative National Private Insurer Policies

http://www.aetna.com/cpb/medical/data/100_199/0105.html

http://www.medsolutions.com/cignaguidelines/guideline_downloads/Chest%20Imaging%20Guidelines-Cigna.pdf

http://apps.humana.com/tad/Tad_New/Search.aspx?criteria=breast+imaging&searchtype=freetext&policyType=both

http://www.unicare.com/medicalpolicies/policies/mp_pw_a053263.htm

https://www.unitedhealthcareonline.com/ccmcontent/ProviderII/UHC/en-US/Assets/ProviderStaticFiles/ProviderStaticFilesPdf/Tools%20and%20Resources/Policies%20and%20Protocols/Medical%20Policies/Medical%20Policies/Breast_Imaging_for_Screening_and_Diagnosing_Cancer.pdf

http://www.medsolutions.com/cignaguidelines/guideline_downloads/Chest%20Imaging%20Guidelines-Cigna.pdf

Aetna, CIGNA (through MedSolutions Inc.), Humana, UniCare, UnitedHealthcare, and Wellpoint/Anthem, cover breast MRI as an adjunct to mammography for patients with familial history of breast cancer, genetic abnormalities, or in those who received radiation therapy to the chest between the ages of 10 and 30. Additionally, Aetna, UniCare, and WellPoint/Anthem cover MRI as an adjunct to annual mammography in women with dense breasts and a personal history of breast cancer. Humana and UnitedHealthcare cover breast MRI as an adjunct to mammography when heterogeneous or extremely dense breast tissue is identified, regardless of breast cancer history. CIGNA does not cover MRI for women whose only risk factor is dense breast tissue.

3.3 Digital Breast Tomosynthesis

Medicare

There are no published national or local coverage determinations for DBT.

Representative Regional Private Insurer Policies

The Regence Group

http://blue.regence.com/trgmedpol/docs/cpsrad55.pdf

The Regence Group considers DBT to be incident to either screening or diagnostic mammogram, and does not reimburse additional costs for procedure.

Premera Blue Cross

https://www.premera.com/medicalpolicies/CMI_126318.htm

Premera Blue Cross considers DBT investigational for screening and diagnosis.

Health Net

https://www.healthnet.com/portal/provider/content/iwc/provider/unprotected/working_with_HN/content/medical_policies.action#B

Health Net considers DBT investigational for all indications, and does not currently cover it.

National Private Payers

http://www.aetna.com/cpb/medical/data/500_599/0584.html

http://www.medsolutions.com/cignaguidelines/guideline_downloads/Chest%20Imaging%20Guidelines-Cigna.pdf

https://cignaforhcp.cigna.com/public/content/pdf/coveragePolicies/medical/mm_0123_coveragepositioncriteria_mammography.pdf

http://apps.humana.com/tad/Tad_New/Search.aspx?criteria=breast+imaging&searchtype=freetext&policyType=both

http://www.unicare.com/medicalpolicies/policies/mp_pw_c142751.htm

https://www.unitedhealthcareonline.com/ccmcontent/ProviderII/UHC/en-US/Assets/ProviderStaticFiles/ProviderStaticFilesPdf/Tools%20and%20Resources/Policies%20and%20Protocols/Medical%20Policies/Medical%20Policies/Breast_Imaging_for_Screening_and_Diagnosing_Cancer.pdf

DBT is considered experimental, investigational or unproven for ***any*** purpose by Aetna, CIGNA, Humana, UniCare, UnitedHealthcare and WellPoint/Anthem.

4. Status of Breast Density Legislation

4.1 Washington

A bill was introduced in the Washington State Legislature in January 2014 that would require physicians to include notification of breast density to patients with their mammogram results. The bill was referred to the Senate Committee on Health Care, and was amended to alter the language of the notification letter that patients would receive. The revised notification limited physician liability beyond notification, added information about breast self-exam, and excluded mention of association between dense breast tissue and increased risk for breast cancer, as this claim was thought to be unproven. Notification language included in the revised bill was as follows:

> *"Your mammogram shows that your breast tissue is dense. Dense breast tissue is common but can make it harder to evaluate the results of your mammogram. Adult women of all ages are encouraged to perform a monthly breast self-exam. This information about your mammogram results is given to you to raise your awareness and to inform your conversations with your primary care provider. A report of your results was sent to your primary care provider (S.S.B. 6050. 2014)."*

The amended bill was passed by the committee on Health Care and referred to the Rules Committee, where it also passed. In the House, the bill was referred to the Health Care and Wellness Committee. The bill was referred back to the Senate Rules Committee at the end of the 2014 regular legislative session, and no further progress has been reported (S.B. 6050, 2014).

Supporters of the bill suggested that notification mandates would ensure that women were aware of their breast density and able to discuss the implications with their doctors. Opponents, including the Washington State Medical Association (WSMA), cited a lack of scientific evidence to support the claim that dense breast tissue increases the risk for breast cancer and expressed concern that notification would lead to an increase in unnecessary ultrasounds for patients with dense breasts. Opponents also felt that the language of the notification would create undue feelings of confusion, anxiety, and fear among patients (Senate Bill Report SSB 6050, 2014). The WSMA suggested evaluation of this topic by the Washington Health Technology Assessment program to further examine the role of dense breast tissue in breast cancer (WSMA, 2014).

4.2 Status of Legislation in Other States

As of October 2014, 19 states have passed legislation mandating breast density notification. These mandates are currently operational in Alabama, California, Maryland, New York, Texas, Virginia, North Carolina, Tennessee, Minnesota, Hawaii, Arizona, Nevada, Rhode Island, Connecticut, Pennsylvania, Oregon, and New Jersey, and will take effect in both Massachusetts and Missouri in 2015.

In addition to density notification, Connecticut and New Jersey have also passed bills requiring insurance coverage of supplemental screening. Illinois and Indiana have existing mandates for insurance coverage and are currently working on bills to mandate density notification. The mandates in Illinois and Connecticut require coverage of ultrasound in women with dense breasts. The breast density law in Indiana requires insurance plans to provide coverage of "appropriate medical screening, tests, or examinations of women with dense breasts," without mention of specific modalities (Ind. Code § 25-22.5-13.2. 2013). New Jersey's insurance mandate requires coverage of ultrasound, MRI, or DBT as supplemental screening in women classified as having extremely dense breast tissue (State of New

Jersey 215th Legislature, 2013). Legislation for mandatory coverage of supplemental screening tests is pending in Pennsylvania.

States that are currently developing or considering notification bills include Maine, South Carolina, Florida, Ohio, Indiana, Michigan, Illinois, Iowa, North Dakota, Colorado, Delaware, and Washington.

4.3 Nationwide Breast Density Legislation

The federal Mammography Quality Standards Act (MQSA) was first enacted in 1994 to ensure that all mammography facilities maintain uniform quality standards. The FDA, which is responsible for enforcing MQSA standards, has acknowledged that changes in mammography technology processes have occurred over time, and that the language in the MQSA may be worth revisiting. As part of this process, a recommendation was made to standardize the reporting of mammographic breast density nationwide. In spring of 2014, the FDA proposed regulatory amendments to the MQSA to account for these changes and recommendations. A notice of proposed rulemaking has been set for December 2014 (Federal Register, 2014). Separately, the Breast Density and Mammography Reporting Act (H.R. 1302, 2011) was introduced in the U.S. House of Representatives in 2011. The act would require all mammography facilities to inform patients with mammographically-dense breast tissue about breast density, the association of density with breast cancer risk and masking, and the possible benefits of supplemental screening. The bill was referred to the House Energy and Commerce Committee (Subcommittee on Health), but was never brought to the full House for a vote. The original co-sponsor of the bill, Rep. Rosa DeLauro (D-CT), reintroduced the bill in October 2013, where it was referred back to committee. In July of 2014, Rep. DeLauro again introduced the bill to the House, where it was referred to the House Energy and Commerce Committee (Subcommittee on Health). Also in July of 2014, Senators Dianne Feinstein (D-CA) and Kelly Ayotte (R-NH) introduced an identical bill in the U.S. Senate, titled the Breast Density and Mammography Reporting Act of 2014. At last report, the bill was referred to the Senate Committee on Health, Education, Labor, and Pensions (Library of Congress, 2014).

5. Previous Health Technology Assessments and Systematic Reviews

We were able to identify three health technology assessments evaluating the use of adding digital breast tomosynthesis to mammography as a screening or diagnostic tool, and one focusing on the use of advanced imaging following negative mammograms in women with dense breasts. We also found two systematic reviews – one on the use of HHUS alone and one the use of both HHUS and ABUS – assessing the use of ultrasound for adjunctive screening in women at average risk for breast cancer. Because of MRI's importance, we have also summarized the major assessments of the use of MRI in women at high risk for breast cancer.

5.1 Formal Health Technology Assessments of Digital Breast Tomosynthesis (DBT)

Blue Cross BlueShield Association Technology Evaluation Center (BCBS TEC, 2014):
http://www.bcbs.com/blueresources/tec/vols/28/28_06.pdf

In a technology assessment on the use of digital breast tomosynthesis with mammography for breast cancer screening or diagnosis, available data suggests that the evidence supporting the addition of breast tomosynthesis is weaker for women who are recalled for diagnostic workup than in the general screening population. For screening purposes, adding digital breast tomosynthesis is more accurate (and possibly more sensitive) than mammography alone. The available evidence also suggests that adding digital breast tomosynthesis to diagnostic mammography in women recalled for screening can reduce the number of unnecessary biopsies. There is insufficient evidence on adding digital breast tomosynthesis to mammography regarding the effect on health outcomes for both screening and diagnostic purposes.

Canadian Agency for Drugs and Technologies in Health (CADTH, 2013):
http://www.cadth.ca/media/pdf/htis/oct-2013/RC0482-Tomosynthesis-Final.pdf

A literature search revealed no evidence on the effectiveness of tomosynthesis used alone compared to digital mammography, and limited evidence on the combination compared to digital mammography for screening or diagnosis. Tomosynthesis plus digital mammography detected more cancers and reduced recall rates in a screening population, but superiority was not conclusive for diagnostic purposes. Women younger than 50 with dense breasts may see the greatest reduction in recall rates with use of this advanced imaging technology.

Institut National d'Excellence en Santé et en Services Sociaux (INESSS, 2014)
http://www.inesss.qc.ca/fileadmin/doc/INESSS/Rapports/Oncologie/INESSS_Tomosynthese_mammaire_num%C3%A9rique.pdf

A technology assessment conducted by a public agency in Quebec identified 22 publications evaluating the use of DBT as both a screening and diagnostic tool in asymptomatic women and found that DBT would simultaneously increase the cancer detection rate and decrease unnecessary recalls. Although concerns were raised over the lack of large, population-based studies, the available evidence suggests that DBT has the potential to be integrated into a structured screening program, pending additional studies confirming its advantage in subgroups.

5.2 Formal Health Technology Assessments of Supplemental Screening in Women with Dense Breasts

Blue Cross BlueShield Association Technology Evaluation Center (BCBS TEC, 2014):
http://www.bcbs.com/blueresources/tec/vols/28/28_15.pdf

In a technology assessment on screening asymptomatic women with dense breasts and a normal mammogram, available data suggests that digital mammography is more sensitive than film mammography. The combination of mammography and ultrasound is more sensitive than mammography alone, but also results in more false positives and unnecessary biopsies. Evidence also suggests that MRI is more sensitive than mammography, although this evidence was generated in women with dense breasts who were also at high cancer risk from other factors. There is insufficient evidence on automated breast ultrasound and DBT in women with dense breasts.

California Technology Assessment Forum (CTAF, 2013):
http://www.ctaf.org/assessments/supplemental-cancer-screening-women-dense-breasts

In an evidence review focused on supplemental screening in women with dense breast tissue and a normal mammogram, a majority of CTAF concluded that the evidence is adequate to demonstrate that supplemental screening with any technology provides more benefit than harm in women at high overall risk of breast cancer (i.e., 5-year risk >3%). However, CTAF found the evidence to be inadequate to suggest that the benefits of supplemental screening in women at low (<1.7%) or moderate (1.7-3%) 5-year risk outweigh the harms. When asked to rank the screening modalities in terms of the preferred choice in high-risk women, CTAF ranked MRI first, followed by HHUS, ABUS, and DBT. CTAF Panel members also voted that the evidence is adequate to demonstrate that digital mammography is superior to film mammography for women with dense breast tissue, and that compared to film mammography, digital mammography greatly reduces the risk of "masking" of breast cancers.

In comparisons of value relative to the lowest-cost test available (HHUS), CTAF considered both ABUS and MRI to be of reasonable value. DBT was felt to be of low value due to limited evidence of effectiveness as a supplemental screening tool and increased radiation exposure.

Canadian Agency for Drugs and Technologies in Health (CADTH, 2007):
http://www.cadth.ca/media/pdf/I3010_MRI-Breast-Cancer_tr_e.pdf

Based on available evidence on screening women at high risk (no RCTs available), MRI was found to be more sensitive and cost-effective compared to mammography. The high risk category included women who were BRCA1 or BRCA2 carriers, their first degree relatives, and those with a strong family history of breast cancer. There was no mention of breast density.

Health Information and Quality Authority, Ireland (HIQA, 2013):

http://www.hiqa.ie/healthcare/health-technology-assessment/assessments/surveillance-of-women-under-50-with-increased-risk-of-breast-cancer

In women age <50 with an elevated risk of breast cancer, evidence suggests that a combination of MRI and digital mammography or MRI alone is more sensitive but less specific than mammography alone. MRI also contributes to decreased breast cancer but at an increased cost. Nevertheless, they estimated that offering annual MRI to women age 30-49 at moderate (10 year risk of 3-8%) or high (10-year risk

>8%) risk would be cost-saving relative to an ad-hoc (i.e., at clinician discretion) surveillance approach or no surveillance.

New England Comparative Effectiveness Public Advisory Council (CEPAC, 2013):

http://cepac.icer-review.org/wp-content/uploads/2013/11/CEPAC-Final-Report-Supplemental-Breast-Cancer-Screening_Master.pdf

In an evidence review focused on supplemental screening in women with dense breast tissue and a normal mammogram, a majority of CEPAC concluded that the evidence is adequate to demonstrate that supplemental screening with any technology provides more benefit than harm in women at high overall risk of breast cancer (i.e., 5-year risk >3%). CEPAC found the evidence to be inadequate to suggest that the benefits of supplemental screening in women at low (<1.7%) 5-year risk outweigh the harms, but adequate for women at moderate (1.7-3%) 5-year risk. When asked to rank the screening modalities in terms of the preferred choice in high-risk women, CEPAC ranked MRI first, followed by DBT; no member voted for ABUS or HHUS as their first choice recommendation. CEPAC Panel members also voted unanimously that the evidence is adequate to demonstrate that digital mammography is superior to film mammography for women with dense breast tissue, and that compared to film mammography, digital mammography greatly reduces the risk of "masking" of breast cancers.

In comparisons of value relative to the lowest-cost test available (HHUS), CEPAC considered both HHUS and MRI to be of reasonable value. CEPAC Panel members abstained from voting on the value of DBT or ABUS compared to HHUS because of insufficient evidence on their clinical effectiveness relative to other screening modalities.

New Zealand Health Technology Assessment Program (NZHTA, 2007):

http://www.otago.ac.nz/christchurch/otago014084.pdf

In women at high risk of breast cancer, ultrasound has equivalent sensitivity to mammography but produces more false positives. The high risk category includes women with a strong family history of breast cancer including women with and without known genetic mutations which predispose to breast cancer. There was no mention of breast density. As with mammography, sensitivity with ultrasound decreases as risk status increases. MRI-based surveillance is more sensitive than mammography, ultrasound, or mammography and ultrasound combined, as accuracy does not appear to diminish with increased risk status. MRI does produce more false-positives than mammography, although these decreased over time in studies evaluating whether a "learning curve" was present for MRI interpretation.

Ontario Health Technology Assessment (2010):

http://www.health.gov.on.ca/english/providers/program/mas/tech/reviews/pdf/rev_%20breast_cancer_screening_20100316.pdf

Data suggests that digital mammography is significantly more sensitive than film mammography in women with heterogeneously or extremely dense breast tissue of any age, asymptomatic women age <50, and those who are pre- or peri-menopausal. There is no evidence of differences in recall rates, however. Evidence suggests that the sensitivity of MRI is significantly higher than that of film mammography in women at high breast cancer risk due to genetic and/or familial factors, regardless of age. There is moderate evidence to suggest that the combination of mammography and MRI is significantly more sensitive than either modality alone in women at high risk of breast cancer from genetic/familial factors, although specificity is either unchanged or decreases.

5.3 Systematic Reviews of Digital Breast Tomosynthesis (DBT)

Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. *Breast.* Apr 2013;22(2):101-108.

http://www.sciencedirect.com/science/article/pii/S0960977613000192

Houssami and Skaane published a systematic review that summarized the results of 14 studies of DBT. None of the studies were randomized trials, addressed breast density subgroups, or reported follow-up data on breast cancer specific mortality or distant-disease recurrence. They report that the studies are preliminary, but suggest that DBT has promise in reducing both false positives and false negatives relative to or when added to digital mammography. The authors highlight five large trials that are ongoing in population-based breast cancer screening programs internationally. They conclude that at this time there is insufficient evidence to justify the widespread use of DBT.

5.4 Systematic Reviews of Magnetic Resonance Imaging (MRI) of the Breast for Women at High Risk

Lord 2007

Lord SJ, Lei W, Craft P, 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.* Sep 2007;43(13):1905-1917.

http://www.sciencedirect.com/science/article/pii/S0959804907004844

Lord and colleagues published a systematic review and meta-analysis of 5 studies that evaluated the impact of MRI added to mammography in young women at high risk of breast cancer.[136] They found consistent evidence that MRI was more sensitive (93-100%) than mammography (25-59%). The addition of MRI to women with negative mammograms identified an additional 10 to 24 cancers per 1,000 examinations. There was an increase in false positive results (71-74 additional false positive follow-ups per 1,000 screens, a 3 to 5-fold increase) and an increase in biopsies with a benign diagnosis (7-46 additional benign biopsies per 1,000 screens). There were no studies assessing whether the addition of MRI reduces patient mortality, interval, or advanced breast cancer rates. They conclude that the benefits of MRI in young, high-risk women have not been established by the evidence.

Warner 2008

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. *Annals of Internal Medicine*. May 6 2008;148(9):671-679.

http://annals.org/article.aspx?articleid=740814

Warner and colleagues published a systematic review and meta-analysis of 11 studies that evaluated the impact of MRI added to mammography in women at high risk of breast cancer.[137] Using a BI-RADS assessment of 4 or 5 as the definition of a positive test, they found consistent evidence that adding MRI to mammography increased the sensitivity for breast cancer from 32% to 84% with a corresponding decrease in specificity from 98.5% to 95.2%. The PPV3 of MRI added to mammography was 25.0%. The authors conclude that annual screening with MRI and mammography is the most accurate approach to screening women with strong familial or genetic predisposition to breast cancer.

5.5 Systematic Reviews of Hand-held Ultrasonography (HHUS) of the Breast

Cochrane review

Gartlehner G, Thaler K, Chapman A, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. *The Cochrane database of systematic reviews.* 2013;4:CD009632.

http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD009632.pub2/abstract;jsessionid=1495F3C42D5F3ED564A5977F83106936.d04t03

The 2013 Cochrane review of the use of ultrasound in addition to mammography to screen average risk women for breast cancer was recently published.[138] The authors concluded, "No methodologically sound evidence is available justifying the routine use of ultrasonography as an adjunct screening tool in women at average risk for breast cancer." For women with dense breasts they concluded, "despite the increased risk for breast cancer and the limitations of mammography in women with dense breast tissue, the available evidence supporting the use of adjunct ultrasonography as a screening tool in women with dense breasts (BI-RADS 3-4) is limited and has to be interpreted cautiously."[138]

Nothacker 2009

Nothacker M, Duda V, Hahn M, 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.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760575/

Nothacker and colleagues performed a systematic review of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue.[139] They did not identify any randomized trials or prior systematic reviews. They identified six cohort studies[140-145] of fair quality, but only two of the studies reported follow-up[141,142] and both were inadequate. They estimated that the cancer detection rate with supplemental ultrasound was 3.2 per 1,000 screens among women with negative mammograms and BI-RADS 2 – 4 density (scattered fibroglandular density through extremely dense). They concluded that there is limited evidence that an additional ultrasound examination after a negative mammogram is useful for the detection of breast cancer in women with mammographically dense tissue.

Scheel 2014

Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts [published online ahead of print June 21, 2014]. *Am J Obstet Gynecol*. doi: 10.1016/j.ajog.2014.

http://www.ajog.org/article/S0002-9378(14)00628-0/abstract

Scheel and colleagues systematically reviewed the literature on supplemental screening ultrasound in women with dense breasts. They evaluated 12 studies, including 10 using hand-hand ultrasound and two on automated breast ultrasound, and found that adjunctive screening with either modality both improved cancer detection and increased the number of false-positive biopsies, with no significant difference between the two technologies. The authors estimated the median additional cancer detection rate for women with dense breasts at 4.2 per 1,000 examinations, but also noted that this number would likely be closer to 2 per 1,000 in routine clinical practice. Most of the lesions identified by ultrasound were small, early-stage cancers, some of which may not have progressed before detection

with routine mammography screening. They concluded that the evidence on supplemental screening with ultrasound is limited, and should not be recommended for women dense breasts alone.

5.6 Systematic Reviews of Automated Whole Breast Ultrasonography (ABUS)

No systematic reviews or technology assessments were found for ABUS.

6. Ongoing Clinical Trials

Information on ongoing clinical studies that have been submitted to the U.S. National Institutes of Health's registry of publicly- and privately-supported studies (www.clinicaltrials.gov) is presented in the table below and on the following pages. We focused on randomized controlled trials and comparative cohort studies.

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Magnetic Resonance Imaging (MRI)					
Breast Cancer Screening With MRI in Women Aged 50-75 Years With Extremely Dense Breast Tissue: the DENSE Trial (Phase 4) NCT01315015	RCT	MRI (n=7,237) DM (n=28,948)	n = 36,185 Age 49 - 75yrs Density > 75% (D4) Females only Negative mammographic assessment (1 or 2)	Number of interval cancers between the MRI group and the control group	December 2019
Familial MRI Screening Study (FaMRIsc) NTR2789	RCT	MRI+CBE (n=1,000) DM+CBE (n=1,000)	Ages 30-55 years Lifetime risk 20% to 49% Exclude BRCA1/2	Number and stage of screen detected cancers stratified by breast density. Secondary: false positive rate, sensitivity, PPV	January 2015

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Contrast-enhanced MR Imaging as a Breast Cancer Screening in Women at Intermediate Risk (MRIB) NCT02210546	RCT	MRI (n= 1,000) DM + US (n= 1,000)	n = 2,000 Age 40 – 59 years Females only 10-year risk 5% or lifetime risk 15-30% Density > 75% No signs, symptoms or previous diagnosis of breast cancer No previous cancer diagnosis No BRCA, p53 genetic mutation No breast implants No hormonal therapy for ovarian function enhancement in past 3 years No life-threatening disease	Rate of in situ, invasive breast cancer detected in either arm Sensitivity, specificity, predictive value of MRI	May 2016
Hand Held Ultrasound (HHUS)					
Japan Strategic Anti-cancer Randomized Trial (J-START) UMIN000000757	RCT	DM + HHUS (n=50,000) DM (n=50,000)	Ages 40-49 years Exclude prior breast cancer	Sensitivity and specificity, incremental cancer detection rate.	March 2016

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Ultrasound and Mammography for Screening Breast Cancer in Chinese Women NCT01880853	RCT	DM HHUS DM+HHUS	n = 47,709 Ages 30-65 years Females only Not pregnant or lactating No breast implants No metastatic disease No symptoms	Cancer detection, sensitivity, specificity, PPV, NPV	December 2011 (no update provided)
Automated Whole Breast Ultrasound (ABUS)					
Earlier Breast Cancer Detection Using Automated Whole Breast Ultrasound With Mammography, Including Cost Comparisons NCT00649337	RCT	ABUS DM	n = 4650 Age: 35 – 90yrs Females only No screening mammogram in the past 10 months No history of breast cancer for at least one year	Number of breast cancers detected one year after screening: ABUS vs blinded DM	January 2010 (no update provided)

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Digital Breast Tomosynthesis (DBT)					
Comparison of Diagnostic Performance of Digital Breast Tomosynthesis (DBT) and Ultrasound (US) in Women With Dense Breasts NCT01910103	Cohort	DBT HHUS	n = 825 Age > 20 Density >50% Females only No previous history of breast surgery or breast core biopsy performed within the prior 6 months	Performance of DBT and US in detecting breast cancer in screening and diagnostic settings	April 2014 (currently listed as recruiting participants)
Comparison of Full-Field Digital Mammography With Digital Breast Tomography for Screening Call-Back Rates NCT01236781	NonRCT	DM Combination of 2-D and 3-D DBT	n = 550 Age > 25 No history of breast cancer Females only n=500 asymptomatic; scheduled for DM n=50 patients with positive DM screenings 30 days prior to registration Not pregnant or lactating Exclude women with breasts too large to allow for adequate positioning for the DBT examination No breast implants	Recall rates (1 year) between DM and DBT	June 2012 (no update provided)

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Digital Breast Tomosynthesis Versus Digital Mammography: A National Multicenter Trial NCT01524029	Cohort	FFDM DBT	n = 600 Age > 20 (> 40 for screening cohort) Females only No breast cancer history Asymptomatic, referred for early detection screening Not pregnant or breast feeding No breast implants	Specificity of DBT in breast lesion characterization Secondary: sensitivity of DBT in malignant breast lesion detection.	September 2012 (no update posted)
Malmö Breast Tomosynthesis Screening Trial (MBTST) NCT01091545	Cohort	DM + DBT	N = 15,000 Ages 40-74 years Females only Not pregnant	Incremental cancer detection rate, sensitivity, and specificity	Primary completion date: June 2014 Study completion date: March 2016

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
Other					
Automated Breast Ultrasound and Digital Breast Tomosynthesis Screening Compared to Full Field Digital Mammography in Women With Dense Breasts NCT02042456	Cohort	DBT + DM DBT + ABUS + DM	n = 650 Age ≥ 18 years Moderate to high risk of breast cancer Not screened in past year No breast implants No pregnant, lactating women No breast cancer, signs of breast cancer in past year	Abnormal interpretation rate both overall and for each group	June 2017

Title/ Trial Sponsor	Study Design	Comparators	Patient Population	Primary Outcomes	Estimated Completion Date
A Multi Modality Surveillance Program for Women At High Risk for Breast Cancer NCT00989638	Cohort	DM MRI Breast biopsy	n = 500 Age > 18 Females only High risk for cancer including dense breasts, known BRCA 1 or 2 mutations, other hereditary breast cancer susceptibility genes No active cancer at enrollment Not pregnant No breast surgery within two weeks of study entry No previous bilateral mastectomy (prophylactic or therapeutic) No history of kidney disease or abnormal kidney tests	Early detection of small or pre-cancerous lesion(s) using a combination of screening measures including biomarkers	June 2015

7. Methods

Objectives

The primary objectives of the systematic review were to:

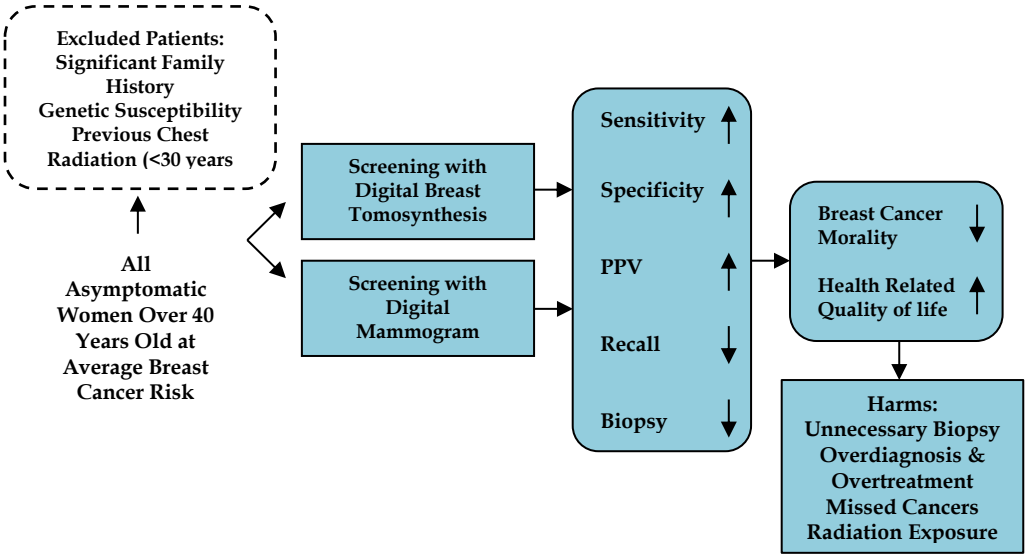
1. Evaluate and compare the published evidence on the effectiveness of screening with digital breast tomosynthesis (DBT) vs. digital mammography, expressed in terms of reduced mortality, increased cancer detection, reduced recalls and unnecessary biopsies, improved health-related quality of life, and other outcomes of interest;
2. Evaluate and compare the published evidence on the effectiveness of supplemental screening with handheld ultrasound (HHUS), automated breast ultrasound (ABUS), and magnetic resonance imaging (MRI) in women with dense breast tissue and a negative mammographic or DBT result, expressed in terms of reduced mortality, increased cancer detection, improved health-related quality of life, and other outcomes of interest;
3. Evaluate and compare the harms of these imaging tests, including overdiagnosis and overtreatment, unnecessary biopsy as a result of false-positive imaging, patient anxiety, and radiation exposure;
4. Examine the differential effectiveness and safety of the tests of interest according to such factors as age, race or ethnicity, comorbidities, BMI, method of breast density classification, overall breast cancer risk, scan vendor, and imaging protocol (e.g., whether ultrasound is performed by a radiologist, technologist, or some combination of the two);
5. Assess the published evidence on costs and cost-effectiveness (e.g., cost per cancer detected) of the imaging modalities of interest; and
6. Evaluate the costs, cost-effectiveness, and budgetary impact to the state of Washington of implementing the imaging strategies of interest.

Analytic Framework

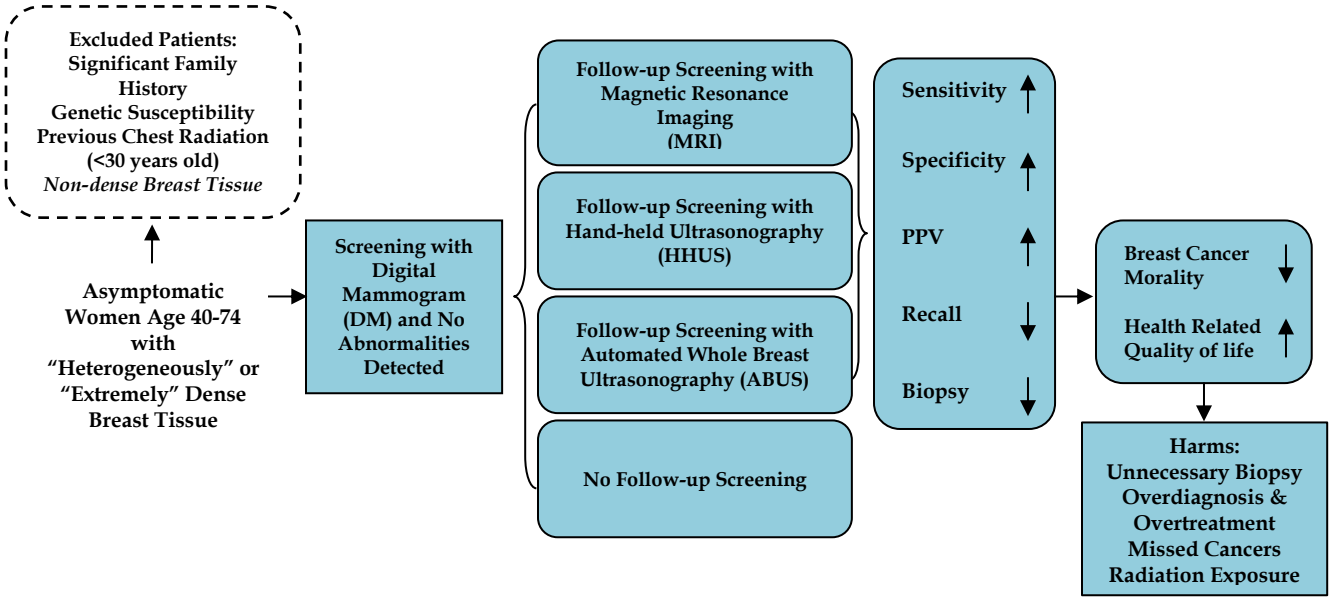
Two distinct analytic frameworks were used for this review, as shown in the Figure on the following page. "Search A" relates to the use of DBT or digital mammography in in all women eligible for screening, and "Search B" depicts the conceptual flow in consideration of the evidence for the supplemental screening tests of interest among women with a negative mammogram or negative DBT result. Note that the figure is intended to convey the conceptual links involved in evaluating outcomes of the imaging modalities of interest, and is not intended to depict a clinical pathway through which all patients would flow. As is the case for many screening or diagnostic tests, it was expected that data linking screening modalities to direct patient outcomes will be limited, requiring instead a series of conceptual links between test characteristics and the major outcomes of interest.

Analytic Framework: Breast Cancer Screening

Search A



Search B



The available literature varies with respect to how directly the impact of these imaging modalities are measured. Some studies are randomized or observational comparisons focused directly on survival, health-related quality of life, and long-term harms, while in other studies a series of conceptual links must be made between test performance and longer-term benefits and/or harms.

Patient Populations

As described above, the population of interest for the assessment of DBT included all asymptomatic women age 40-74 who are candidates for screening mammography every 1-2 years. Studies were stratified by screening interval. The target population for the comparison of supplemental screening modalities included women with dense breast tissue and a normal mammography or DBT result. We examined clinical trials and observational studies that included women in the BI-RADS categories of "c" (heterogeneously dense) or "d" (extremely dense) (BI-RADS, 2013). Both populations were stratified by a number of important characteristics as the available evidence allowed, including age, race/ethnicity, overall breast cancer risk, and others. Given that the method of breast density classification has changed over time and may vary between institutions, we also recorded as many details on how this classification was applied as possible.

Interventions

We evaluated the effectiveness, costs, and cost-effectiveness of magnetic resonance imaging (MRI), handheld ultrasonography (HHUS), automated ultrasonography (ABUS), and digital breast tomosynthesis (DBT). Data on these technologies were collected regardless of manufacturer, imaging protocol, or other test characteristics. Note that, while the focus of attention on supplemental screening technologies was on findings in women with dense breast tissue, overall results from major clinical studies were also abstracted to provide context for test performance.

Comparators

The primary comparator of interest for frontline screening with DBT was digital mammography, as nearly 95% (12,790/13,523) of all US mammography machines accredited by the U.S. Food and Drug Administration are now full-field digital (FDA, 2014). However, given that the change-over from film mammography to digital was relatively recent, and film may in fact still be a standard in certain geographies, we did not exclude any study comparing DBT to film mammography. We evaluated supplemental screening technologies against each other, and individually against additional follow-up (with any method) or no follow-up examination in women with dense breasts. In addition, we considered studies utilizing clinical breast examinations (CBEs) or self-exams as a comparator.

Outcomes

Specific outcomes of interest included the test characteristics of the modalities of interest, including rates of sensitivity and specificity, positive predictive value, recall, and biopsy. We obtained data on cases of cancer detected, cancers missed, including "interval cancers" (I.e., those that were detected in the interval between screening tests). Where available, we also collected data on the impact of screening modality on breast cancer mortality and health-related quality of life. Finally, potential harms of interest included unnecessary biopsy as a result of false-positive imaging, overdiagnosis and overtreatment, missed cancers, and radiation exposure.

Information on the costs and cost-effectiveness of each screening method was also collected where available.

Timeframe

Data on all relevant measures were abstracted at all relevant timepoints, regardless of study duration.

Study Designs

Data from both RCTs and selected types of observational studies were considered for measures of effectiveness. Observational studies of interest included those making explicit prospective or retrospective comparisons of each screening modality of interest to another screening method within the same setting as well as comparisons across settings.

No limits were placed on study selection based on sample size, duration, location, or frequency of outcome measurement. We do note, however, that studies that are less than 12 months' duration cannot reliably measure test performance because they cannot capture all interval cancers. Nevertheless, we included these studies if information on other outcomes of interest such as rates of cancer detection, recall, and/or biopsy were reported.

Literature Search and Retrieval

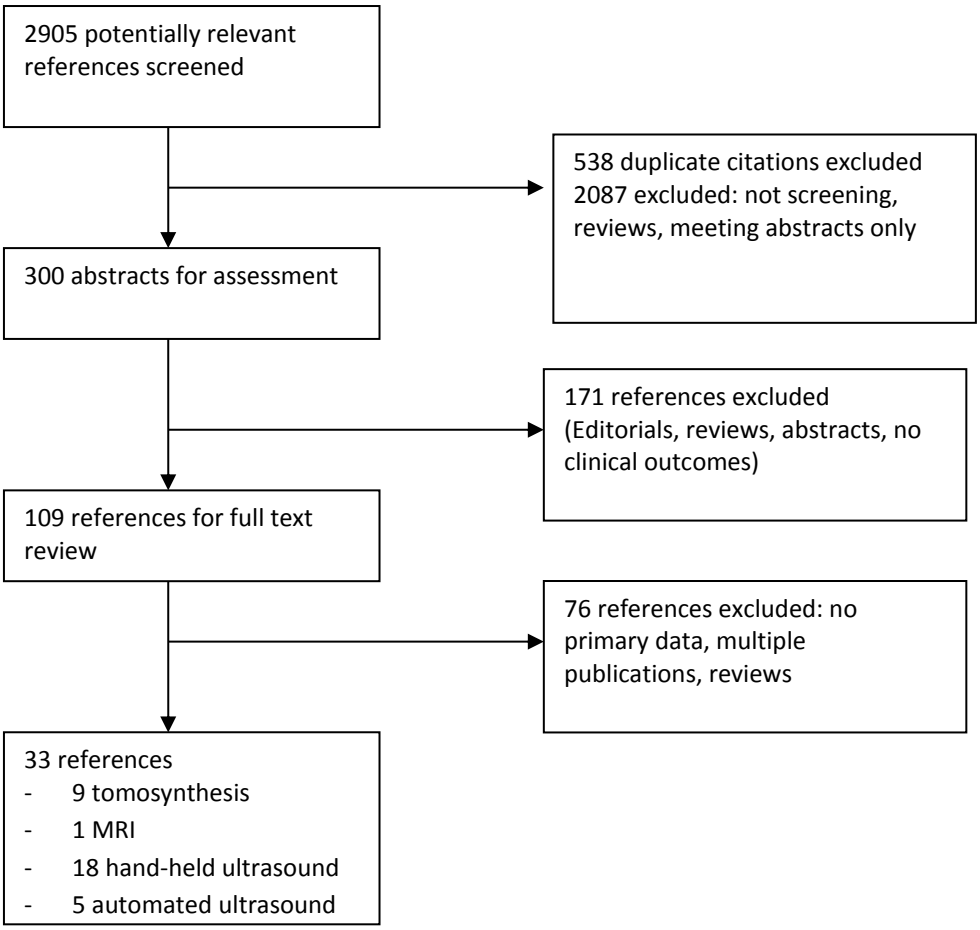
The general timeframe for literature search and retrieval was January 1990 – November 2014. We focused on English-language reports only. Publications that appeared after the search period but prior to submittal of the final report were considered as well.

The electronic databases we searched as part of the systematic review included MEDLINE, EMBASE, and *The Cochrane Library* (including the Database of Abstracts of Reviews of Effects [DARE]) for health technology assessments (HTAs), systematic reviews, and primary studies. Reference lists of all eligible studies were also searched and cross-referenced against public comments received by the HCA. The strategies used for MEDLINE, EMBASE, and *The Cochrane Library* are shown in Appendix A.

Studies of DBT required a comparator but were not otherwise restricted in terms of entry criteria. For evaluations of supplemental screening tests, we included all studies of imaging technologies used to screen for breast cancer in women with a recent (within 30 days) negative digital mammogram and high breast density. A negative mammogram was defined by BI-RADS assessments 1 or 2 (negative or benign finding) and high breast density was defined as heterogeneously dense ("c") or extremely dense ("d") using the BI-RADS density criteria. As mentioned above, if digital mammography was not used, we included film mammography for completeness.

The combined search results identified 2,905 potentially relevant studies for this assessment (Figure 1 on the following page). After elimination of duplicate and non-relevant references, we identified a total of 33 studies for inclusion. Nine studies evaluated the use of DBT in women undergoing routine screening. Nine studies evaluated the use of DBT in women undergoing routine screening. Of the remaining 24 studies examining supplemental screening in women with dense breast tissue and a negative mammogram, we identified 18, five, and one for HHUS, ABUS, and MRI respectively. The primary reasons for study exclusion were (a) focus on high-risk populations (BRCA carriers, lifetime risk > 25%, personal history of breast cancer, recent diagnosis of breast cancer); (b) use of imaging for diagnosis of a suspicious mass or abnormality only; (c) studies were only reader reliability studies or descriptions of ongoing studies without results; or (d) non-dense breasts for the supplemental screening search. Because the initial search criteria excluded many of the studies initially identified, we expanded our inclusion criteria to capture studies that defined density as scattered fibroglandular densities, heterogeneously dense or extremely dense tissue and those that only reported cancer detection rates. We also included a summary of the American College of Radiology Imaging Network (ACRIN) 6666 study because it is widely cited and relevant to practice patterns in the United States. Finally, to give some perspective on the relative value of MRI, we summarized the larger studies and systematic reviews of MRI to screen women at very high risk due to hereditary breast cancer or its equivalent.

Figure 1. PRISMA flow chart showing results of literature search.



The two most important outcomes in breast cancer are breast cancer specific mortality and disease-free survival. Because early stage breast cancer has such a long natural history and the majority of women do well, large randomized trials with long follow-up are needed to demonstrate the improvements in these outcomes in patients screened with additional imaging technologies. For short-term studies, the potential benefit of additional screening is best summarized by the incremental cancer detection rate. The potential harms can be assessed by evaluations of the false positive rate, the recall rate, the biopsy rate, the positive predictive values, and the radiation dose.

Study Quality

We used criteria published by the U.S. Preventive Services Task Force to assess the quality of RCTs and comparative cohort studies, using the categories “good”, “fair”, or “poor”. Guidance for quality rating using these criteria is presented below (AHRQ, 2008).

Good: *Meets all criteria: Comparable groups are assembled initially and maintained throughout the study (follow-up at least 80 percent); reliable and valid measurement instruments are used and applied*

equally to the groups; interventions are spelled out clearly; all important outcomes are considered; and appropriate attention to confounders in analysis. In addition, for RCTs, intention to treat analysis is used.

Fair: *Studies will be graded "fair" if any or all of the following problems occur, without the fatal flaws noted in the "poor" category below: Generally comparable groups are assembled initially but some question remains whether some (although not major) differences occurred with follow-up; measurement instruments are acceptable (although not the best) and generally applied equally; some but not all important outcomes are considered; and some but not all potential confounders are accounted for. Intention to treat analysis is done for RCTs.*

Poor: *Studies will be graded "poor" if any of the following fatal flaws exists: Groups assembled initially are not close to being comparable or maintained throughout the study; unreliable or invalid measurement instruments are used or not applied at all equally among groups (including not masking outcome assessment); and key confounders are given little or no attention. For RCTs, intention to treat analysis is lacking.*

In addition, the QUADAS-2 tool was used specifically to measure the quality of diagnostic accuracy studies (Whiting, 2011). Overall strength of evidence for each screening modality of interest was described as “high”, “moderate”, or “low”, and utilized the evidence domains employed in the AHRQ approach (AHRQ, 2012). The quality of the supplemental screening studies was evaluated using an adaptation of the QUADAS-2 specific to concerns regarding supplemental screening (see Table 6 on pages 44 through 50). Specifically, patient inclusion criteria were expanded to consider (a) the use of digital vs. film mammography as the initial test; and (b) the approach used to classify breast density (some studies accepted BI-RADS category 2 or “b” [scattered densities]). In keeping with standards set by the Washington HCA, however, assignment of strength of evidence will focus primarily on study quality, quantity of available studies, and consistency of findings.

Finally, summary ratings of the comparative clinical effectiveness and comparative value of the tests of interest (i.e., *across* multiple key questions) will be assigned using ICER’s integrated evidence rating matrix (Ollendorf, 2010). The matrix has been employed in previous Washington HCA assessments of virtual colonoscopy, coronary CT angiography, cervical spinal fusion for degenerative disc disease, cardiac nuclear imaging, and most recently, proton beam therapy.

Data from all retrieved studies were included in evidence tables regardless of study quality. However, the focus of attention in presentation of results was primarily on good- or fair-quality studies.

8. Results

8.1 Evidence Quality

The quality of available evidence on: (a) use of digital breast tomosynthesis versus digital mammography as a frontline general population screening tool; and (b) use of automated and handheld ultrasound as well as magnetic resonance imaging for supplemental screening in women with dense breast tissue, is summarized in the tables on pages 44 through 51. We identified a total of 33 studies. Importantly, *none* of the studies identified assessed the impact of either general population or supplemental screening on breast cancer morbidity or mortality. Nine studies evaluated the use of DBT in women undergoing routine screening. Of the remaining 23 studies examining supplemental screening in women with dense breast tissue and a negative mammogram, we identified 17, five, and one for HHUS, ABUS, and MRI respectively. Most of the comparative studies identified had major quality concerns. The majority of studies were observational in nature; no randomized control trials were available. Because the population of interest differed for each search, we used separate criteria to assess study quality.

All of the nine studies evaluating the use of DBT were rated "poor" based on the QUADAS-2 assessment, primarily because of no or incomplete follow-up; breast cancer screening studies should be at least as long as the interval between screening tests (i.e., 1-2 years) to adequately capture interval cancers. Only one study (Destounis, 2014) had adequate follow-up for evaluating interval cancers; however, patient groups were imbalanced with respect to breast cancer risk factors and other clinical characteristics in this study. While all studies included asymptomatic women presenting for routine screening, limited availability of DBT at screening sites and selection and/or volunteer bias (i.e., systematic differences between patients who agree to undergo a test and the target population for study) were issues in six of the studies (Destounis, 2014; Freidewald, 2014; Greenberg, 2014; Haas, 2013; McCarthy, 2014; Rose, 2013). The only two prospective studies (Ciatto, 2013; Skaane, 2013) had adequate reporting of nearly all outcomes of interest; number of biopsies as a result of abnormal findings was not included in either study, however. Study quality is summarized in Table 5 page 55.

The quality of the supplemental screening studies was evaluated using an adaptation of the QUADAS-2 specific to concerns regarding supplemental screening. The single study (Kuhl, 2010) evaluating the use of MRI in women with dense breasts was found to be of fair quality, as consecutive women were screened using the various MRI protocols, and two-year follow-up data were recorded in nearly all women (98%). Nevertheless, this was not a true screening study as nearly half of women had a personal history of breast cancer; in addition, applicability to U.S. settings is limited, as the majority of women had not only a negative mammogram but a negative ultrasound result as well.

Although we identified the most studies on the use of HHUS as a supplemental screening tool in women with dense breasts, only one study (Berg, 2012) was judged to be of good quality. The interval between the mammogram and the HHUS examination should be relatively short (i.e., with one month of each other). Otherwise the HHUS may find cancers that would also be visible on a mammogram at a later point in time. One study (Hooley 2012) included HHUS results from as much as 361 days after the mammogram – it is likely that mammography performed at that point would find additional cancers as well. Twelve studies did not report the time interval between examinations, one study reported that

there was an average of two months between the examinations, and three studies performed both examinations within the same month.

Four of the five studies evaluating ABUS as a supplemental screening tool were of poor quality. There was a high degree of uncertainty for several outcomes, including biopsy and recall rate. The former was not reported in most studies, and the latter was reported with significant variation across studies. It is likely that these inconsistencies are in part due to the high degree of technical proficiency required to perform the procedure, though only one study (Arleo, 2014) addressed this issue.

Table 5: Quality assessment of DBT studies.

Study	Interval Between DM and DBT	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
Ciatto 2013	Same day	Yes, 95% agreed to participate.	No, no interval cancers	None reported	Prospective, consecutive	Digital	Poor
Destounis 2014	Same day	No, selection bias	Yes	122 (23%) in DM group, 77 (15%) in DM + DBT lost to follow-up	Retrospective	Digital	Poor
Friedewald 2014	Same day	No, depending on availability	No, no interval cancers	N/A (not patient-level)	Retrospective pre - post	Digital	Poor
Greenberg 2014	Same day	No, volunteer bias	No, no interval cancers	208 lost to follow-up after recall, 73 lost after biopsy recommendation	Retrospective	Digital	Poor
Haas 2013	Same day	No, selection bias	No, no interval cancers	None reported	Retrospective	Digital	Poor
Lorenco 2014	Same day	Yes, all patients screened with DBT after implementation	No, no interval cancers	11% lost to 1 year follow-up; 1% withdrew	Retrospective pre - post	Digital	Poor
Lorenco 2014	Same day	Yes, all patients screened with DBT after implementation	No, no interval cancers	11% lost to 1 year follow-up; 1% withdrew	Retrospective pre - post	Digital	Poor
Rose 2013	Same day	No, volunteer bias	No, no interval cancers	None reported	Retrospective, pre - post	Digital	Poor

Study	Interval Between DM and DBT	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
Skaane 2013	Same day	Yes, 70% agreed to participate	Incomplete follow-up for interval cancers	None reported	Prospective, consecutive	Digital	Poor

* Quality rating (based on QUADAS-2 domains of patient selection, index testing, reference standard, and study flow/timing):

Good - consecutive sample from asymptomatic women presenting for population-based screening, no or few withdrawals, follow-up sufficient to capture all outcomes of interest

Fair – same as above, but with some risk of bias or applicability concerns based on patient selection or loss to follow-up

Poor – any combination of (a) insufficient follow-up for outcome determination; (b) substantial selection bias; (c) substantial and/or differential loss to follow-up; or (d) inappropriate or inadequately reported interval between index test and reference standard

Table 6: Quality assessment of supplemental screening studies.

Study	Interval Between Tests	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
MRI							
Kuhl 2014 Germany	Same day	Yes, referred on clinical grounds	Yes	8 lost at 1 year follow-up	Prospective	Digital	Fair
HHUS							
Maestro 1998	NR	Unclear. 19% with personal history of BC.	No, incomplete follow-up.	None reported	Unclear	Film	Poor
Buchberger 2000	NR	Not reported	No, no follow-up	None reported	Unclear	Film	Poor

Study	Interval Between Tests	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
Kaplan 2001	NR	No. Some with palpable or focal abnormal mammographic findings in other quadrants included.	Yes	6 recommended for biopsy	Prospective	Film	Fair
Kolb 2002	NR	NR, but appears to be consecutive.	No, incomplete follow-up.	None reported	Unclear	Film	Poor
Korpraphong 2014	Same day	No, volunteer bias	No, no interval cancers	None reported	Unclear	Digital	Poor
Crystal 2003	NR	NR	No, no follow-up	None reported	Unclear	Film	Poor
Leconte 2003	NR	No	No, no follow-up	None reported	Unclear	Film	Poor
Brancato 2007	Within 1 month	Unclear: only 20.3% of eligible enrolled	No, no interval cancers	None reported	Unclear	Film	Poor
de Felice 2007	Same day	Yes, though no description of the participants age, family history, etc.	No, no interval cancers	None reported	Prospective	Film	Poor
Corsetti 2008	~50% same day	Yes	No, no interval	None reported	Retrospective	Film	Poor

Study	Interval Between Tests	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
	~50% within 4 weeks		cancers				
Corsetti 2011	NR	Yes	Yes	None reported	Retrospective	Film	Fair
Hooley 2012	Mean 61 days Range 0-361 days	No: included BI-RADS 0 assessment on mammogram	Yes	17% did not return for one year follow-up	Retrospective	Digital	Poor
Leong 2012	NR	Yes	Yes	28% of negatives with no follow-up	Prospective	Digital	Fair
Weigert 2012	NR	Only 30% of eligible participated	No, not all interval cancers.	11/429 recommended for biopsy	Retrospective	NR	Poor
Chae 2013	Same day	Only 40% of eligible participated; also offered regardless of mammography results	Yes	None reported	Retrospective	Digital	Poor
Girardi 2013	NR	Yes	No, no interval cancers	None reported	Retrospective	Digital	Poor

Study	Interval Between Tests	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
MRI							
Kuhl 2014 Germany	Same day	Yes, referred on clinical grounds	Yes	8 lost at 1 year follow-up	Prospective	Digital	Fair
HHUS							
Maestro 1998	NR	Unclear. 19% with personal history of BC.	No, incomplete follow-up.	None reported	Unclear	Film	Poor
Buchberger 2000	NR	Not reported	No, no follow-up	None reported	Unclear	Film	Poor
Kaplan 2001	NR	No. Some with palpable or focal abnormal mammographic findings in other quadrants included.	Yes	6 recommended for biopsy	Prospective	Film	Fair
Kolb 2002	NR	NR, but appears to be consecutive.	No, incomplete follow-up.	None reported	Unclear	Film	Poor
Korpraphong 2014	Same day	No, volunteer bias	No, no interval cancers	None reported	Unclear	Digital	Poor
Crystal 2003	NR	NR	No, no follow-up	None reported	Unclear	Film	Poor

Study	Interval Between Tests	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
Leconte 2003	NR	No	No, no follow-up	None reported	Unclear	Film	Poor
Brancato 2007	Within 1 month	Unclear: only 20.3% of eligible enrolled	No, no interval cancers	None reported	Unclear	Film	Poor
de Felice 2007	Same day	Yes, though no description of the participants age, family history, etc	No, no interval cancers	None reported	Prospective	Film	Poor
Corsetti 2008	~50% same day ~50% within 4 weeks	Yes	No, no interval cancers	None reported	Retrospective	Film	Poor
Corsetti 2011	NR	Yes	Yes	None reported	Retrospective	Film	Fair
Hooley 2012	Mean 61 days Range 0-361 days	No: included BI-RADS 0 assessment on mammogram	Yes	17% did not return for one year follow-up	Retrospective	Digital	Poor
Leong 2012	NR	Yes	Yes	28% of negatives with no follow-up	Prospective	Digital	Fair
Weigert 2012	NR	Only 30% of eligible participated	No, not all interval cancers.	11/429 recommended for biopsy	Retrospective	NR	Poor

Study	Interval Between DM and DBT	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
Chae 2013	Same day	Only 40% of eligible participated; also offered regardless of mammography results	Yes	None reported	Retrospective	Digital	Poor
Girardi 2013	NR	Yes	No, no interval cancers	None reported	Retrospective	Digital	Poor
Parris 2013	NR	No, abnormal mammograms included and 11% not dense	No, no interval cancers.	None reported	Retrospective	Digital	Poor
Venturini 2013	NR	No, additional screening tailored to individual risk and breast density	No, only women with BI-RADS 3 lesions had follow-up	None reported	Prospective	Digital	Poor
Berg 2012 ACRIN 6666	< 91 days	No – high-risk	Yes	<10%	Prospective	Mix film and digital	Good, but wrong population and not 100% digital
ABUS							
Arleo 2014	Same day	NR	No, no follow-up	None reported	Retrospective	NR	Poor

Study	Interval Between DM and DBT	Representative Spectrum – Consecutive Patients For Screening Exam	Appropriate Reference Standard	Withdrawals	Design	Mammo	Quality*
Kelly 2010	468 women with ABUS 6 months from mammogram	No. Only 5% participation at some sites; up to 25% at others. 22% diagnostic. Data incomplete for high density subgroup.	Yes	Unclear: only 80% had mammographic follow-up > 1 year after initial imaging.	Prospective	1/3 digital, 2/3 film	Poor
Brem 2014	Same day	Yes	No, no interval cancers	None reported	Prospective	Digital	Fair
Stoblen 2011	Same day	Yes	No, no interval cancers	None reported	Prospective	Digital	Poor
Giuliano 2013	NR	NR	Yes	None reported	Prospective	Digital	Poor

* Quality rating:

Good - consecutive sample from women presenting for screening, digital mammography with a negative assessment, BI-RADS 3 or 4 density, supplemental screening test done within one month of mammogram, at least 90% follow-up of benign and negative findings at one year, and complete reporting of "positive" and "negative" results for the supplemental screening test;

Fair – same as above but film mammography and BI-RADS 2-4 density acceptable;

Poor – less than 90% follow-up or a non-consecutive sample (spectrum bias) or supplemental screening test done more than one month after the mammogram or incomplete reporting of positive and negative results for advanced imaging

Key Question #1: What is the effectiveness of screening with digital breast tomosynthesis (DBT) vs. digital mammography among women aged 40-74 who are candidates for screening mammography?

8.2 Digital Breast Tomosynthesis

We identified a total of nine studies (Ciatto, 2013; Skaane, 2013a; Skaane, 2013b; Haas, 2013; Rose, 2013; Friedewald, 2014; Destounis, 2014; Lourenco, 2014; Greenberg, 2014; McCarthy, 2014) evaluating the use of DBT in comparison to digital mammography. The primary data are summarized in Table 7 below. Two large prospective studies (Skaane, 2013; Ciatto, 2013) compared the test characteristics of digital mammography with and without DBT performed in the same patients on the same day. Three additional studies study (Destounis, 2013; Haas, 2013; Greenberg, 2014) compared two groups of patients, one screened with digital mammography alone and the other with digital mammography plus DBT. Finally, two retrospective, multicenter studies (Friedewald, 2014; Rose, 2013) and two single-center studies (Lourenco, 2014; McCarthy, 2014) used data from prior screening examinations before and after implementation of DBT. Table 7 below summarizes the nine DBT studies that compare the use of DBT to digital mammography. These studies include asymptomatic women presenting for routine screening in various sites across the United States and Europe. Some findings of interest are not reported, as follow-up was of inadequate duration or is currently incomplete to capture information on interval cancers.

Table 7: Studies comparing DBT to digital mammography for screening of asymptomatic women.

		----------DBT----------					Sensitivity		Specificity	
Study	Women, N	Recall rate/1,000	CDR /1,000	Bx rate /1,000	PPV1 %	PPV3 %	M %	DBT %	M %	DBT %
Ciatto 2013	7,292	42.9	8.1	NR	18.8	NR	66.1	100	95.5	96.6
Skaane 2013	12,621	61.1	8.0	NR	13.1	NR	62.6	82.1	93.8	94.6
Haas 2013	6,100	84.0	5.7	NR	6.8	NR	100	100	NR	NR
Friedewald 2014	173,663	91.0	5.4	19.3	6.4	29.2	NR	NR	NR	NR
Rose 2013	9,499	54.5	5.4	10.6	9.8	24.7	100	100	91.7	95.1
Destounis 2014	524	42.0	5.7	11.5	13.6	50	100	75	97.9	99.4
Lourenco 2014	25,498	93.4	5.4	17.4	6.2	23.8	NR	NR	91.1	94.0
Greenberg 2014	59,617	135.8	6.3	26.3	4.6	23.8	NR	NR	84.3	87.0
McCarthy 2014	18,220	103.7	5.5	17.7	6.2	25.4	NR	NR	NR	NR

M: Mammography; DBT: Digital Breast Tomosynthesis

CDR = cancer detection rate; Bx rate = biopsy rate PPV1 = positive predictive value of a positive test result; PPV3 = positive predictive value of biopsies actually performed

Skaane and colleagues (Skaane, 2013a; Skaane, 2013b) recently published initial results from a large series of patients evaluated with both digital mammography and DBT performed on the same day. The study evaluated DBT in 12,621 women presenting for routine screening mammography in Oslo, Norway in 2011. DBT added an average of 10 seconds per view to the time required for mammography (40 seconds total). The reading time also increased, from 45 seconds for mammography to 91 seconds for mammography plus DBT. The total radiation dose increased from 1.58 mGy for digital mammography to 1.95 mGy for DBT (note: this was prior to the introduction of software allowing for creating of mammographic images during DBT scanning). According to standard practice in Norway, two radiologists independently interpreted the images for each woman and the potentially positive cases were reviewed at an arbitration meeting. Follow-up is not complete, but three interval cancers were identified during nine months of follow-up. These were not included in the statistics reported in the paper, but they have been counted as false negatives in the calculations performed for this assessment.

Addition of DBT decreased both false positives and false negatives. Thus DBT had higher sensitivity (82.1% compared to 62.6%) and specificity (94.6% compared to 93.8%) than digital mammography alone (both p<.001). The cancer detection rate increased from 6.1 to 8.0 cases per 1,000 examinations (p=.001) while the recall rate decreased from 67.2 to 61.1 per 1,000 (p<.001). The adjusted increase in cancer detection was 40% (Rate Radio [RR] 1.40, 95% CI 1.13 to 1.71, p<.001).

There are several issues that make it difficult to generalize the results of this study to the U.S. The standard of care in Norway is to have two radiologists interpret each mammogram and to have an arbitration meeting to review all positive results and decide which to call back. As noted earlier, this approach has a much lower recall rate than that observed in the United States (Hofvind, 2008). In addition, follow-up for interval cancers was incomplete.

A similar study (Ciatto, 2013) compared digital mammography to DBT in 7,292 women coming in for routine screening mammography at two population-based centers in Italy. As in Norway, two radiologists independently interpreted the images for each woman. However, if either was positive, the woman was recalled in this study while in Norway, there was a conference to decide who should be recalled. The Italians found that DBT had greater sensitivity (100% compared to 66.1%) and greater specificity (96.5% compared to 95.5%) compared to digital mammography (both p<.0001). This translated into a higher cancer detection rate (8.1 compared to 5.3 per 1,000 examinations, p<.0001) with a lower recall rate (42.9 compared to 49.5 per 1,000 examinations). As in the prior study (Skaane, 2013), there was no long-term follow-up, so the primary outcomes were the cancer detection rate and the recall rate. The investigators also did not report the biopsy rate in the study.

The primary concern with the Ciatto study is the lack of follow-up for interval cancers. This artificially raises the sensitivity of DBT to 100% and causes an overestimation of the specificity and negative predictive value as well. A recently published post-hoc analysis of the Ciatto study (Houssami, 2014) comparing outcomes for different screening strategies found six additional cancers after the first year of follow-up and estimated the interval cancer rate to be 0.82 cancers per 1,000 screens (95% CI: 0.30–1.79/1,000). However, because all participants received integrated mammography and DBT, it was not possible to determine the independent impact of DBT in reducing interval cancers, and the authors suggested that the interval cancer rate from this analysis be interpreted with caution.

In a recently published analysis (Destounis, 2014) of U.S.-based DBT experience, Destounis and colleagues compared results among patients choosing to undergo DBT plus digital mammography (n=524) to those from a sample set of randomly selected women (n=524) screened with digital mammography alone during the same timeframe (June - December 2011) at a facility in New York. The

combination of DBT and digital mammography had a significantly lower recall rate (42.0 compared to 114.5 per 1,000 examinations, p<.0001) compared to digital mammography alone. DBT was also associated with a lower biopsy rate (11.5 compared to 22.9 per 1,000 examinations), a higher cancer detection rate (5.7 compared to 3.8 per 1,000 examinations), and a higher positive predictive value for those undergoing biopsy (50.0% [3/6] compared to 16.7% [2/12]); these differences did not appear to be statistically tested, however. The population examined appeared to be very low risk, as only six cancers were detected among 1,048 women screened (0.6%). After one year of follow-up, two women in the digital mammography group had a cancer diagnosis, neither of which were interval cancers. There were four cancers detected in the DBT group, of which one was an interval cancer. However, nearly 20% of women in the study did not complete one year follow-up, so these results are incomplete at best. In addition, selection bias could not be ruled out, as no adjustments were made for differences in breast cancer risk between groups, and the DBT group had higher proportions of women with dense breast tissue, personal or family history of breast cancer, and atypia on prior biopsy.

Another retrospective comparative cohort study conducted at four sites in the U.S. was released online by Haas and colleagues on July 30, 2013 (Haas, 2013). They compared the recall rate and cancer detection rate at sites using digital mammography (n=7,058) to those at sites using DBT (n=6,100). All women presenting for screening mammography were included except those with breast implants or those with large breasts requiring tiled images. As in the prior studies, DBT decreased the recall rate (84 compared to 128 per 1,000 examinations, p<.01) but the cancer detection rate was not significantly increased (5.7 compared to 5.2 per 1,000 examinations, p=.70).

This retrospective study (Haas, 2013) also has several major methodological concerns. As with the Destounis study, the mammography and DBT groups were not well matched. Women in the DBT group were younger (55.8 years compared to 57.5 years), had more extremely dense breasts (5.6% versus 3.0%) and less fatty breasts (8.8% versus 13.8%), were more likely to have a personal history of breast cancer (5.5% versus 2.8%), and were more likely to have a first-degree relative with breast cancer (18.8% versus 15.9%). The investigators did not adjust for these differences in their primary analyses, but did present the results of logistic regression analyses adjusted for age, breast density, family history and personal history of breast cancer. In those analyses DBT was associated with a 35% reduction in the odds of recall (p<.0001). The investigators did not report biopsy rates, so it is not possible to determine whether the reduction in the recall rate translated into a similar reduction in breast biopsies. Finally, there was no follow-up for interval cancers so the sensitivity, negative predictive value, and specificity cannot be calculated.

The largest DBT study conducted to date was a retrospective multicenter study (Friedewald, 2014) that evaluated the screening performance of 13 U.S. academic and nonacademic breast cancer screening centers over two periods – one year prior to the introduction of DBT, and one year following. After adjusting for site differences, the investigators retrospectively evaluated a total of 454,850 examinations, of which 281,187 were screened with digital mammography alone, and 173,663 with a combination of digital mammography plus DBT. While the number of recalls was lower for DBT compared to digital mammography (91 versus 107 per 1,000 screens, p<.001), the biopsy rate was higher for DBT (19.3 versus 18.1 per 1,000 screens, p=.004). Twelve of the 13 sites increased cancer detection with DBT with an overall rate of 5.4 compared to 4.2 per 1,000 examinations for mammography alone (p<.001). The addition of DBT was also associated with a significantly higher PPV for recalls (6.4% compared to 4.3%, p<.001) and for biopsy (29.2% compared to 24.2%, p<.001).

Despite its size and impressive findings, this study nevertheless carries some significant limitations beyond those of its pre-post design. As with the prospective studies described above, the Friedewald study lacks long-term follow-up for interval cancers, and therefore cannot provide a full picture of test performance. In addition, clinical information was limited to data required for reporting to regulatory authorities, so there may have been heterogeneity in the two populations analyzed that could not be accounted for. Finally, statistical calculations were based on screen-level rather than patient-level data so it was not possible to consider each record to be an independent observation, which limited the statistical adjustments that could be made.

A newly-published retrospective multicenter study (Greenberg, 2014) compared 38,674 patients who underwent digital mammography to 20,943 patients who voluntarily selected DBT over a 16-month period. Although the study site description is not definitive, it appears that at least one of the sites also participated in the Friedewald study, so there may be some data overlap.

This study found a significant reduction in the recall rate with DBT (135.8 vs. 161.5 per 1,000 examinations), which represents a 2.6% absolute reduction in recalls (p<.0001). The cancer detection rate also increased from 4.9 in the DM group to 6.3 per 1,000 examinations in the DBT cohort, representing a 29% overall increase (p=.035). As with the Friedewald analysis, this study also reported a higher biopsy rate for DBT (26.3 vs. 21.6 per 1,000 examinations, p=.0003), and a PPV3 value that was similar for both groups. Although there were no differences in baseline characteristics between the two groups with regard to age, family history of breast cancer, or breast density, the investigators did find that significantly fewer DBT patients had less than two mammographic views at recall than those in the DM group (74.1% vs. 51%, p<.001), which may suggest that DBT is more effective at identifying patients who require additional follow-up.

Although this was one of two retrospective DBT studies to evaluate contemporaneous screenings, there are some limitations that may impact its findings. First, not only was there the potential for volunteer bias in the selection of DBT, most women were required to pay a $50 premium for the procedure, which may have introduced additional bias in the selection of candidates for DBT. In addition, this study assessed patients immediately following implementation of DBT at various times in several sites across three states; there were no adjustments to account for a learning curve with the new technology or potential heterogeneity within the study population.

A new publication of a retrospective study that also involved one of the participating sites in the Friedewald study (McCarthy, 2014) used a pre-post design to assess 26,299 screenings in 18,220 asymptomatic women over a 17-month period, and found a reduction in the recall rate with DBT that was statistically-significant in both unadjusted and adjusted models (8.8% vs. 10.4% for DM, OR 0.80, 95% CI 0.74-0.88, p<.001). Although this study was not powered to detect significant differences in biopsy rates (1.8% vs. 2.0% for DBT) or cancer detection rates (4.6 vs. 5.5 per 1,000 examinations for DBT) in the overall population, for women under age 50 there was a significant increase in the number of cancers detected with DBT compared to digital mammography (5.7 vs. 2.2 per 1,000 examinations, p=.02).

Another recent retrospective comparative cohort study (Lourenco, 2014) assessed 25,498 patients over one year before and after the implementation of DBT at a breast imaging center. In line with some other recent studies of DBT, biopsies were recommended more frequently in the DBT group (17.4 vs. 16.3 per 1,000 examinations for DM). There were also 31% fewer recalls with DBT (6.4% vs. 9.3%, p<.00001). It is unclear why absolute rates of recall were lower in this study than in others, given that it was U.S.-based. In contrast to recall findings, neither PPV3 nor the cancer detection rate differed

significantly between groups. Finally, as with most of the other studies in this evaluation, follow-up was insufficient to detect interval cancers.

The final study of DBT (Rose, 2013) also used a pre-post design rather than a direct comparison of the two technologies. The investigators compared the screening benchmarks for the combination of DBT and digital mammography (n=9,499) to those of digital mammography alone at the same sites in Texas during the prior year (n=13,856). There was no follow-up for interval cancers, so the sensitivity, specificity and negative predictive value are overestimated. As in the prior studies, DBT had a lower recall rate (54.5 compared to 87.2 per 1,000 examinations, p<.001). The biopsy rate was also lower (10.6 compared to 15.2 per 1,000 examinations) and cancer detection rate was higher (5.4 compared to 4.0 per 1,000 examinations), but these differences were not statistically significant.

Subgroup Analyses: Dense Breast Tissue

DBT is of clinical interest in part because its imaging approach is felt to be more precise when breast tissue is dense. Several of the previously described studies also included subgroup analyses focused on women with dense breasts. The Ciatto study (Ciatto, 2013) published detailed results for the two highest BI-RADS density categories and noted that differences in recall rates were greatest in the dense breast subgroups. The overall difference in recalls was modest (approximately 43 vs. 50 per 1,000 examinations for DBT versus mammography). However, among women with heterogeneously dense breasts (n=4,242) the recall rate was 102 per 1,000 examinations in the DBT group compared to 167 per 1,000 in the digital mammography group (p<.01). Among women with extremely dense breasts (n=555) the recall rate was 67 per 1,000 examinations in the DBT group compared to 156 per 1,000 in the mammography group (p<.01). In contrast, the incremental cancer detection rate for integrated DBT plus digital mammography was similar in the low-density and high-density groups (2.8 per 1,000 compared to 2.5 per 1,000 examinations). The Norwegian study (Skaane, 2013) similarly found that the improvement in the cancer detection rate with DBT was comparable in both the dense and non-dense groups, but did not report on any other outcomes.

In the Rose study (Rose, 2013), a subgroup of women with BI-RADS 3 or 4 density showed similar improvements with DBT across all breast density categories; overall, DBT had a lower recall rate cancer (68.8 compared to 102.8 per 1,000 examinations) and a higher cancer detection rate (5.4 compared to 3.9 per 1,000 examinations), though these results were not tested statistically.

Similarly, the McCarthy study (McCarthy, 2013) did not find any significant differences with regard to breast density for any outcome. The recall rate was significantly lower with DBT compared to DM for both nondense (p=.001) and dense breasts (p=.006), and both groups had comparable outcomes for cancer detection and positive predictive values.

Finally, Haas and colleagues (Haas, 2013) reported that the recall rate reduction was greatest in those with heterogeneously dense breasts (45% reduction, 95% CI 34% to 54%, p<.001) and those with extremely dense breasts (60% reduction, 95% CI 30% to 78%, p=.002). The investigators did not calculate the cancer detection rates in the density subgroups, so it is not clear whether the trend towards increased cancer detection applies to the high-density subgroup.

Summary: Screening DBT

Six studies (Ciatto, 2013; Skaane, 2013a; Haas, 2013; Rose, 2013; Friedewald, 2014; Destounis, 2014) of women presenting for routine screening for breast cancer found that DBT increased the cancer

detection rate relative to mammography while decreasing the recall rate and, in all but one study, the biopsy rate. These results were relatively consistent despite the different methods for interpretation used in the three different countries. Findings were also consistent across both the dense and non-dense subgroups in the studies reporting on those subgroups.

However, there were issues of study heterogeneity as well as comparability of screening populations. As shown in Table 8 on the following page, there is substantial uncertainty with recall rates, since the two prospective studies come from outside the U.S. where patterns of recall differ markedly. In addition, test performance is likely to be overstated in all of these studies, as follow-up is not long enough in any one study other than the Destounis study, and 20% of women in that study did not achieve one year of follow-up (Destounis, 2014).

Table 8: Estimated yield of DBT in combination with digital mammography vs. digital mammography alone in women presenting for general population screening.

Statistic	Digital mammography	DBT+Digital mammography	Uncertainty
Recall rate per 1,000	100-160	80-140	Moderate-high
Biopsy rate per 1,000	14-22	12-27	Moderate
CDR per 1,000	3-5	4-6	Moderate-high
PPV3	20-25%	25-30%	Low-moderate

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

Key Question #2: What is the comparative effectiveness of handheld ultrasonography, automated ultrasonography, and magnetic resonance imaging when used as supplemental screening modalities in women with dense breast tissue and a negative mammogram or negative DBT result?

8.3 Screening Magnetic Resonance Imaging (MRI) of the Breast

The search identified only one study that evaluated the benefit of MRI following negative mammography in women with dense breasts. A recently published prospective study (Kuhl, 2014) evaluated the use of an abbreviated protocol (AP) form of MRI, consisting of a pre- and post-contrast acquisition (called MIP and FAST images, respectively) in breast cancer screening to evaluate its ability to detect cancers in asymptomatic women presenting for screening in Germany; findings were compared with a traditional full diagnostic MRI protocol (FDP). Women with heterogeneously dense or extremely dense breasts and a normal or benign mammography result as well as a negative or benign ultrasound (n=427) were included. All women were categorized according to their individual risk of developing breast cancer (i.e., mild, moderate, or high).

Acquisition time for AP was 184 seconds versus 1,024 seconds for FDP; reading time for an MIP image was 2.8 seconds and less than 30 seconds for AP. Both AP and FDP identified 11 cancers representing an additional cancer yield of 18.5 per 1,000 screens with 10 of the 11 cancers being identified using MIP images alone. Specificity (94.3% versus 93.9%) and positive predictive value (24.4% versus 23.4%) were not statistically different between AP and FDP readings, and negative predictive value was 100% for both protocols (95% CI, 99.3-100). No additional cancers were detected in the second screening round.

Although well-matched with regard to study population, the Kuhl study has several limitations. First, MRI was used as a third-line screening in women following a normal mammogram and negative or benign ultrasound result, which does not reflect current practice in the U.S. Nearly half of screened women had a personal history of breast cancer (220/443, 49.6%), so the high cancer detection rate likely reflects this underlying risk in the women referred for supplemental screening. Finally, outcomes from this study may not be applicable to screening in community clinical practice, as the radiologists assigned to review the images were considered expert readers.

Magnetic Resonance imaging (MRI) for Screening High-risk Women

The majority of studies we identified from our search evaluated the use of MRI in women at a high risk of developing breast cancer, without regards to breast density. In particular, several large studies have evaluated the test characteristics of MRI in conjunction with mammography and ultrasound in BRCA1 and BRCA2 mutation carriers and other women at very high risk for breast cancer. Those studies are summarized briefly below as an update to the HCA review of MRI in such women conducted in 2010 (Delfini Group, 2010). The ACRIN 6666 study offered MRI to high-risk women who completed the third round of annual screening ultrasound and mammography in that study. The results will be described in the section on HHUS below.

Magnetic resonance imaging (MRI) has been primarily studied for breast cancer screening in women deemed to be at high risk either by personal history, family history or because they were known carriers of either a BRCA1 or BRCA2 mutation (Berg, 2004; Hagen, 2007; Kriege, 2004; Kuhl, 2005; Leach, 2005; Sardanelli, 2007; Tilanus-Linthrost, 2000; Warner, 2004; Kriege, 2006; Kuhl, 2000; Lehman, 2005; Lehman, 2007; Podo, 2002; Morris, 2003; Port, 2007; Stoutjesdijk, 2001; Yu, 2008; Warner, 2001; Trecate, 2006). These women have a lifetime risk greater than 20%, rather than the 10% to 20% lifetime risk for most women with high breast density. No studies have demonstrated that MRI reduces the risk of death from breast cancer; there are no studies comparing women screened with MRI to other women screened with mammography alone and none of the studies of the test characteristics of MRI are of sufficient duration or size to evaluate patient-oriented outcomes such as the breast cancer recurrence or death from breast cancer.

Table 9 on page 60 summarizes the larger prospective screening studies (n=5,652) that compare the use of MRI in high-risk woman to mammography with or without ultrasound. Women in these studies were primarily BRCA1 or BRCA2 mutation carriers or their first degree relatives. None of the studies followed women for more than one or two years. In addition, the majority of these studies compared MRI to film mammography only, since digital mammography was not widely disseminated until after publication of the DMIST trial in 2005.

The sensitivity of MRI for breast cancer in Table 9 ranged from 77% to 100%. The sensitivity of mammography (25%-59%) and ultrasound (13%-65%) in these studies was about half that of MRI. In the largest three studies (Kriege, 2004; Kuhl, 2005; Leach, 2005), which included 52% of the cancers in all 14 studies, the sensitivity of MRI ranged from 71% to 91% while the sensitivity of mammography ranged from 32% to 40%. However, the specificity of MRI is consistently lower than mammography. In the same three studies, the specificity of MRI ranged from 81% to 97% compared to 93% to 99% for mammography, and in each individual study the specificity of MRI was lower than that of mammography. Because breast cancer is relatively uncommon, even in these high-risk populations, the lower specificity of MRI translates into a much higher number of false positive results. One study (Sardanelli, 2007) suggested that the high false positive rate decreases after the initial MRI. In that

study the rate of false positive results declined from 14% initially to 8.2% on subsequent MRI's, but was still substantially higher than the 4.6% false positive rate for mammography (Kriege, 2006).

The cancer detection rate of MRI ranged from eight to 67 per 1,000 examinations in these studies (see Table 9 on the following page) – much higher than the three to six per 1,000 examinations typically reported in studies of mammography (Table 10 on page 39). This reflects in part the higher sensitivity of MRI and in part the higher incidence of breast cancer in these high-risk women.

Table 9: Prospective studies comparing magnetic resonance imaging, ultrasound, and mammography to screen high-risk women for breast cancer.

Study	Women, N	CDR MRI /1,000	Bx rate /1,000	PPV1 MRI %	PPV2 MRI %	PPV3 MRI %	Sensitivity			Specificity		
							M %	HHUS %	MRI %	M %	HHUS %	MRI %
Tilanus-Linthorst 2000	109	28	46	-	60	60	0*	-	100			
Podo 2002	105	67	86	-	89	89	13	13	100			
Kriege 2004	1909	12	29	-	57	57	40	-	71	95	-	90
Warner 2004	236	30	157	-	46	46	36	33	82	99	96	81
Kuhl 2005	529	36	147	-	50	50	32	40	91	97	91	97
Leach 2005	649	29	-	-	-	25	40	-	77	93	-	81
Lehman 2005	367	8	63	-	17	17	25	-	100	98	-	93
Lehman 2007	171	23	82	-	43	43	33	17	100	91		79
Sardanelli 2007	278	22	90	-	60	60	59	65	94	99	98	98
Kuhl 2010	687	15	34	-	48	48	33	37	93	99	98	98
Berg 2012	612	15	70	-	-	19	31	-	88	92	-	76

M: Mammography; HHUS: Ultrasound; MRI: magnetic resonance imaging

CDR = cancer detection rate

Bx rate: biopsy rate

PPV1 = positive predictive value of a positive test result (BI-RADS assessment 0, 4, or 5)

PPV2 = positive predictive value of a biopsy recommended (BI-RADS assessment 4 or 5)

PPV3 = positive predictive value of biopsies actually performed

However, this higher cancer detection rate comes at a cost: the biopsy rates in the MRI studies in Table 9 range from 29 to 157 biopsies per 1,000 examinations. The biopsy rates are lower in studies of screening mammography (10 to 25 per 1,000 examinations). The PPV3 ranged from 17% to 89%, but the median was 48%, which is a very high yield per biopsy.

It is worth noting in Table 12 on page 46 that the sensitivity of mammography and ultrasound were similar to each other in each of the five studies that report the sensitivity of all three screening technologies. The sensitivity of mammography and ultrasound in these studies is much lower than the sensitivity usually reported for these tests. The low sensitivity is due to the large number of cancers that are found by MRI alone – more than typically appear as interval cancers in the year following a screening examination. This suggests that many of the cancers detected by MRI would not have been diagnosed without MRI for more than one year after the examination. Early detection of cancers that would have become clinically apparent at a later date should translate into a higher cure rate and the need for less aggressive therapies, but some proportion of the cancers detected by MRI are likely to represent overdiagnosis – cancers that never would have become symptomatic in a woman's life.

The two systematic reviews described in Section 5 (Previous Systematic Reviews and Technology Assessments) both found that the addition of MRI significantly increased the sensitivity of screening for breast cancer, but increased false positive results; the effect on breast cancer mortality remained unknown because none of the studies had sufficient follow-up duration to evaluate this endpoint (Lord, 2007; Warner, 2008). In one of the meta-analyses (Warner, 2008), adding MRI to mammography increased the sensitivity from 39% to 94%, but decreased specificity from 94.7% to 77.2%. If the prevalence of breast cancer in a high-risk population is 4.4% (the pooled prevalence across the 14 studies), then adding MRI to mammography in 1,000 women would detect an additional 24 breast cancers (increased from 17 to 41) and an additional 167 women would receive false positive results (increased from 51 to 218).

Summary: Screening MRI of the Breast

There is very limited data evaluating MRI in a general screening population with dense breasts, as well as in populations at intermediate risk (15% to 20% lifetime risk). We identified only one study evaluating MRI as a supplemental screening tool in women with dense breasts and negative initial mammography and ultrasound findings. The results from this study show that an abbreviated version of MRI may hold some promise in detecting the presence of additional cancers without sacrificing levels of sensitivity and specificity associated with the full diagnostic protocol, with the potential for cost savings from the abbreviated protocol. However, this study was conducted in Europe, with screening and referral patterns that are not generalizable to the U.S. setting.

The data from high-risk populations suggest that the addition of MRI would more than double the cancer detection rate (best estimate 2.4-fold increase) with a four-fold increase in the recall rate (best estimate 4.3-fold increase). Estimates based on these data are shown in Table 10 below. There is a high level of uncertainty around these values because of the lack of direct evidence from studies of MRI in women with dense breast tissue and because of the heterogeneity of the findings in the studies of high risk women summarized in Table 10.

Table 10: Estimated incremental yield of MRI after negative digital mammography in women with dense breast tissue.

Statistic	Digital Mammography	Incremental yield with MRI	Uncertainty
Recall rate per 1,000	100-120	100-120	High
Biopsy rate per 1,000	14-22	20-40	High
CDR per 1,000	3-5	3-11	High
PPV3	20-25%	22%-48%	High

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

These estimates suggest that MRI would find substantially more cancers than those found by digital mammography, with a PPV3 between 22% and 48%. There would be approximately 100-120 additional recalls and between 20 and 40 additional biopsies per 1,000 women in order to identify approximately 3-11 additional cancers.

8.4 Screening Hand-held Breast Ultrasound (HHUS)

Eighteen studies of almost 100,000 women screened with HHUS met the search criteria for this assessment and are described in Table A1 in Appendix B, with a quality assessment of these studies presented in Table 6 beginning on page 44 (Buchberger, 2000; Kaplan, 2001; Kolb, 2002; Crystal, 2003; Leconte, 2003; Corsetti, 2008; Kolb, 2007; Maestro, 1998; Corsetti, 2006; Brancato, 2007, De Felice, 2007; Corsetti, 2001; Hooley, 2012; Leong, 2012; Weigert, 2012; Chae, 2013; Girardi, 2013; Parris, 2013; Venturini, 2013; Korpraphong, 2014). In general, all participants in these studies underwent mammography first and those with negative mammograms were subsequently screened by HHUS. One study by Corsetti and colleagues is presented twice in the tables: their 2008 publication (Corsetti, 2008) had a large number of examinations; and their 2011 publication (Corsetti, 2011) included one year follow-up for a subset of the women. Results from the ACRIN 6666 trial (Berg, 2008; Berg, 2012) are also described in the tables, although the study did not meet the inclusion criteria. However, it was the only prospective study in the United States with complete reporting of the data on the combination of mammography and HHUS with one-year follow-up after more than one round of screening.

As shown in Table A1 in Appendix B, the participants in these studies had a mean age usually in the 50s with a broad range (25 to 91 years). Most included asymptomatic women presenting for screening mammography who were found to have dense breasts, although the definition of high density varied somewhat. The majority of the trials were done outside of the United States. Three recent retrospective cohorts (Hooley 2012; Weigert 2012; Parris 2013) described the findings in Connecticut, which was the first state to pass a law requiring breast density notification. These three studies represent the best evidence in the US population for the incremental cancer detection rate with HHUS, although they do not include any data on the interval cancer rate. The two other trials in the U.S. (Kaplan 2001; Kolb 2002) reported results from imaging performed in the year 2000 and earlier. A radiologist performed the HHUS in the majority of the studies. Nine of the studies reported no follow-up on participants, three reported variable follow-up on a subset of patients, and three reported one-year follow-up. This is typical for publications of data from mammography facilities, as they keep records on the follow-up of abnormal tests and cancer detection for quality assurance work, but do not routinely follow patients with normal mammography results to identify interval cancers. This also

means that the sensitivity, specificity, and negative predictive value reported from those studies will overestimate the true values. The diagnostic accuracy test characteristics from these studies are summarized in Appendix B. In these studies, when one participant was diagnosed with more than one cancer or had more than one biopsy, the statistics were reported on a per participant basis rather than per cancer or biopsy. The statistics in Table A2 in Appendix B represent only participants who had a negative mammogram assessment and fell into one of the two high density BI-RADS categories (heterogeneously dense or extremely dense) except for those with separate rows for mammography and mammography plus supplemental screening.

Only five of the studies in the Table A2 (Hooley 2012; Leong 2012; Girardi 2012; Parris 2013; Korpraphong, 2014) compared HHUS to digital mammography. Ten studies compared HHUS to film mammography and one did not report the type of mammography machine used in the study. The ACRIN 6666 Trial used a mix of digital and film mammography (Berg, 2008; Berg, 2012).

Only three of the trials (Kaplan 2001; De Felice 2007; Leong 2012) in women with dense breasts and the ACRIN 6666 trial (Berg, 2008; Berg, 2012) were prospective studies. Prospective studies are more likely to have complete and consistent measurement of the key outcomes because they are defined objectively at the start of the study and collected systematically. It is worth noting in Table A2 that these trials had by far the highest recall rates (>100 recalls per 1,000 examinations). Most of the other studies did not systematically report recalls after ultrasound and often reported the HHUS assessment as positive only if a biopsy was recommended, thus underestimating the true recall rate.

Table A2 summarizes the major diagnostic test results from these studies. Because the majority of the studies did not follow women with negative HHUS assessments for interval cancers, the most relevant statistics to focus on are the recall rate, the biopsy rate, the cancer detection rate, the positive predictive value of positive tests (PPV1) and the positive predictive value of biopsies performed (PPV3). As described in the background section of this report, the recall rate for mammography is typically about 100 per 1,000 examinations, the biopsy rate about 10 per 1,000 examinations, the cancer detection rate about 3.5 to 5 per 1,000 examinations, the PPV1 about 4% and the PPV3 about 25%.

The recall rate for HHUS after normal mammography ranged from 21 to 170 per 1,000 examinations with the median value across the studies of 59, lower than the typical recall rate for mammography described earlier in this report of 100 per 1,000 examinations. In the ACRIN 6666 study (Berg, 2012), HHUS recalled 186 women per 1,000 examinations. As noted above, all of the prospective studies (Kaplan, 2001; De Felice, 2007; Leong, 2012; ACRIN 6666) reported recall rates greater than 100 per 1,000 examinations, so these values are likely to be more accurate.

The biopsy rate for women having HHUS after normal mammography ranged from 12 to 114 per 1,000 examinations with a median of 46. In the ACRIN 6666 study the biopsy rate was 88 per 1,000 examinations the first round and about 61 per 1,000 examinations the third round. The cancer detection rate varied from 0.4 to 14.2 per 1,000 examinations with a median value of 3.2 per 1,000 examinations. In the ACRIN 6666 trial, HHUS detected 5.9 cancers per 1,000 examinations. It is not clear why the biopsy rates vary across such a wide range. Potential explanations include incomplete reporting of cyst aspirations, different thresholds for performing cyst aspirations, operator dependency in performing HHUS, and differences in the proportion of patients undergoing a first time screening HHUS compared to those with prior examinations for comparison.

There was also a wide range of estimates across the studies for the PPV1 (2.0 to 11.7%, median 6.5%) and the PPV3 (3.2 to 18.4%, median 7.1%). The heterogeneity of these results was likely due to a

combination of factors. These include the study design (prospective, retrospective), the use of film or digital mammography, differences in the assessment of mammography across countries, whether a radiologist or a technician performed the HHUS, the level of experience and training of the person performing the HHUS, and differences in the populations studied (age distribution, breast cancer risk factors, time since last mammogram).

The characteristics of the cancers detected by mammography alone and of ultrasound among women with a negative mammogram are described in Table A3 in Appendix B. The table shows that most of the cancers detected by HHUS after negative mammography are small, node negative, early stage cancers. These are the cancers that are potentially curable by early detection before they develop into cancers with a poorer prognosis. Cancers at an early stage also require less aggressive therapy: the patient may be eligible for lumpectomy rather than mastectomy and may not require systemic chemotherapy. Thus early detection may improve both quality and quantity of life. The counter-argument is that some of these early stage cancers may not have progressed much before the next routine screening examination with mammography. Thus, they may ultimately have been detected and cured with mammographic screening alone. In addition, some proportion of these cancers may represent overdiagnosis: the identification of a cancer that would not have ever progressed to cause symptoms prior to the death of that individual woman. The identification of such cancers would lead to unnecessary labeling of the woman as someone who has cancer as well as unnecessary surgery and chemotherapy. The only way to test which of these two competing hypotheses is true would be to perform a randomized trial comparing the two approaches to breast cancer screening.

A large retrospective study of asymptomatic women with dense breasts (n=20,864) presenting for routine screening underwent mammography (n=12,505) or mammography plus HHUS (n=8,359) in Korea over a two-year period (Chae, 2013). Screening with both ultrasonography and mammography increased the cancer detection rate (2.9 compared to 0.5 per 1,000 examinations) but also increased the recall rate (55 per 1,000 compared to 42 per 1,000). Of note, while no details were provided on abnormalities were adjudicated, the low cancer detection rates reported here are reflective of the much lower incidence of breast cancer in Asian countries relative to Western nations.

Of the 24 cancers identified with HHUS, 23 were invasive and one was DCIS. Eleven cancers were identified in the mammography-only group: five were false-negatives which were all subsequently identified with diagnostic ultrasonography, five were DCIS, and one was an invasive cancer. As a result, sensitivity of ultrasound was 100.0% compared to 54.5% for mammography alone (p=.002). PPV1 values were also significantly higher for HHUS (5.3% versus 1.1% for mammography, p<.001). However, false-positives were higher in the HHUS group (5.18%) compared to mammography alone (4.14%) and the PPV2 value was also much lower for HHUS plus mammography (11.1% [24/216] versus 50%, [6/12]) than for mammography alone; neither of these differences were tested for statistical significance.

There are several limitations to this study. While a broad spectrum of breast cancer risk was allowed in this study, ultrasonography was chosen voluntarily, women who had an elevated risk may have disproportionately opted for additional screening, thereby which may have artificially increased the cancer detection rate for ultrasound. In addition, women who opted for HHUS were asked to pay out of pocket for the test, which may explain why only 40% of women with dense breasts chose HHUS. Finally, this study does not fully meet criteria because there was no requirement for a negative mammogram before the decision for supplemental screening with HHUS was made.

Another recent study (Korpraphong, 2014) evaluated of the use of hand-held ultrasound in asymptomatic women with non-fatty breasts (BI-RADS 2, 3, or 4 density) who presented for

mammography screening in Thailand and were subsequently examined with HHUS. Subgroups of women were analyzed according to age and breast density. Of the 14,483 screenings, 115 cancers were detected: 31 with mammography alone, 19 with HHUS alone, and 55 with both HHUS and mammography. The overall incremental cancer detection rate was 1.4 per 1,000 examinations ($p<.0001$). Women with extremely dense breasts and women between 40-59 years old had an incremental cancer detection rate of 2.5 and 2.0 per 1,000 examinations, respectively, which were the highest among all categories; statistical significance was not reported for either of these findings, however. In addition, no outcomes were assessed for biopsy rate, recall rate, or interval cancers. The investigators also warn there is a high incidence of breast cancer in the community where the study was conducted, making it difficult to apply these results to the general population.

American College of Radiology Imaging Network (ACRIN) 6666 Trial

Because the ACRIN 6666 trial (Berg, 2008; Berg, 2012) was the only prospective trial performed in the United States with mostly digital mammography and one year follow-up for multiple screening rounds (good quality), its findings are the most pertinent to the focus of this review and will therefore be described in detail below. The population studied was higher risk than that of a typical screening population, so the biopsy rate, cancer detection rate, and positive predictive values will be higher than those of a screening population. For instance, in the first round the biopsy rate based on mammography was 14.4 per 1,000 examinations, the cancer detection rate was 7.5 per 1,000 examinations, and the PPV3 was 31%, all of which are higher than expected for mammography in a screening population (approximately 10 per 1,000, 5 per 1,000, and 25% respectively).

The ACRIN 6666 trial randomized 2,809 high-risk women to receive both mammography (film or digital) and ultrasound in alternate order (Berg, 2008). High-risk was defined by at least one of the following: a personal history of breast cancer; positive for BRCA1 or BRCA2 mutation; a lifetime risk ≥ 25%, a 5-year risk ≥ 2.5% or ≥ 1.7% with extremely dense breast tissue; prior biopsy with atypical ductal hyperplasia, atypical lobular hyperplasia, lobular carcinoma in situ or atypical papilloma; or prior mantle radiation. The study also required that the women have at least one quadrant of one breast with heterogeneously dense or extremely dense tissue on a prior mammogram. The trial did not meet the inclusion criteria for this assessment for two reasons: the study subjects are high risk rather than a general screening population and were not required to have dense breasts by BI-RADS criteria.

The women were followed for three annual cycles and upon completion of the third cycle, the women were offered additional screening with breast MRI (Berg, 2012). There were 2,659 women with data for analysis after the first year of follow-up. Their median age was 55 years and 93% were white. The primary risk factors for inclusion in the study were a personal history of breast cancer (53%), a lifetime risk ≥ 25% (19%), and a five-year risk ≥ 2.5% (15%). The investigators present the results of mammography alone and for the combination of mammography plus ultrasound, but not for ultrasound alone or the subgroup of women with a negative mammography assessment. When possible, we calculated the incremental results for ultrasound following negative mammography.

In the first screening round, mammography detected 20 cancers (cancer detection rate 7.6 per 1,000 examinations) and ultrasound detected an additional 14 cancers (5.9 per 1,000 examinations) (Berg, 2012). There were two interval cancers so the sensitivity of mammography was 55.6% (20/36) and the sensitivity of ultrasound in women with negative mammograms was 87.5% (14/16). The number of recalls increased from 306 with mammography alone to 707 with mammography plus ultrasound, a 2.3-fold increase in the recall rate (from 115.1 per 1,000 examinations to 265.9 per 1,000). The number of

breast biopsies increased from 65 to 272, a 4.2-fold increase (from 24.4 per, 1,000 examinations to 102.3 per 1,000). The PPV3 for ultrasound in women with negative mammograms was only 6.8%.

By the third screening examination, the test characteristics changed, reflecting a reduction in prevalent cancers due to early detection, the transition to digital mammography, and improved specificity with increased experience of the radiologists and the availability of prior examinations available for review (Berg, 2012). Mammography detected 23 cancers (cancer detection rate 9.9 per 1,000 examinations) and ultrasound detected an additional nine cancers (4.2 per 1,000 examinations). There were 14 interval cancers so the sensitivity of mammography was 50.0% (23/46) and the sensitivity of ultrasound in women with negative mammograms was 39.1% (9/23). The investigators did not report the recall rate and biopsy rate for round three, but did report the numbers for the combination of rounds two and three. The number of recalls increased from 453 with mammography alone to 809 with mammography plus ultrasound, a 1.8-fold increase in the recall rate (from 94.1 per 1,000 examinations to 168.1 per 1,000). The number of breast biopsies increased from 97 to 339, a 3.5-fold increase (from 20.1 per 1,000 examinations to 70.4 per 1,000). The PPV3 for ultrasound in women with negative mammograms was 7.1%.

In round three, women were offered MRI in addition to HHUS and mammography (Berg, 2012). The 612 women in the MRI sub-study had higher risk for breast cancer and were younger than those who declined participation (Berg, 2010). In this group of participants, mammography alone detected five cancers, ultrasound detected an additional two cancers (sensitivity for the combination 43.8%, cancer detection rate 11.4 per 1,000 examinations) and MRI detected nine additional cancers (sensitivity 100%, incremental cancer detection rate 14.7 per 1,000 examinations and combined cancer detection rate 26.1 per 1,000 examinations). The nine cancers detected by MRI only were small (median 8.5 mm) and all were lymph node negative. Both cancers seen only with HHUS (not mammography) were also diagnosed with MRI. The high cancer detection rate in the women in the MRI group reflects the high underlying risk for cancer in the women who agreed to participate in the sub-study. The recall rate was 85.0 per 1,000 examinations for mammography alone, 163.4 per 1,000 for the combination of mammography plus HHUS and 260.0 per 1,000 for MRI. The biopsy rate was 62.1 per 1,000 examinations for the combination of mammography plus HHUS and 132.3 per 1,000 for the combination with MRI. The PPV3 for MRI in women with a negative mammogram was 22.4%, which is much higher than that of ultrasound.

In this high-risk population, the ACRIN 6666 study found that supplemental screening with HHUS produced a relatively high yield of cancers the first round of screening, approximately doubling the cancer detection rate, but this decreased with subsequent rounds. In order to find these cancers, the recall rate more than doubled so that one in four women (26.6%) were recalled in the first round. The number of biopsies performed increased by a factor of four. In the first round, the combination of ultrasound plus mammography led to almost as many biopsies (10.2% of women) as women recalled with mammography alone (11.5% of women). The addition of MRI more than doubled the cancer detection rate of mammography plus ultrasound, but was associated with even an even higher recall rate and a doubling of the biopsy rate. The PPV3 for ultrasound in women with negative mammograms was very low (6.8% round 1, 7.1% rounds 2 and 3) compared to mammography alone (29.1% round one, 38.1% rounds two and three). The PPV3 for MRI in women with negative mammograms was 22.4%.

Summary: Screening HHUS of the Breast

There are no studies evaluating the impact of adding HHUS to mammographic screening among women with dense breast tissue that address the key patient-centered outcomes of breast cancer mortality and disease-free survival. The available body of evidence, focusing largely on shorter-term recall rates, biopsy rates, cancer detection rates and false positive rates, is limited by multiple factors. There were a large number of studies, but the heterogeneity of the study designs, populations, and results preclude the use of meta-analytic techniques to combine the results. The majority of the studies used film mammography, were retrospective, did not fully report the recall rate, and were not able to calculate sensitivity because women with negative mammograms were not followed for interval cancer. There is not even one prospectively designed study with one-year follow-up of HHUS in women with a negative mammogram and heterogeneously dense or extremely dense breasts. The best estimates for sensitivity and specificity come from the ACRIN 6666 trial (87.5% and 81.9% respectively) because it is the highest quality study and sensitivity and specificity are usually not influenced by the risk of the population being studied (Berg, 2012). The best estimate for the incremental cancer detection rate is centered around 3-4 cancers per 1,000 examinations, but the results from the three studies on the Connecticut experience were closer to two cancers per 1,000. The results from Connecticut are more likely to be representative of routine clinical practice in the U.S. The recall rates and PPV1's in these studies were greater than those of mammography, indicating that the addition of HHUS approximately more than doubles the recall rate. The recall rate doubled in the ACRIN 6666 study as well. Finally, the biopsy rates were three to five times higher than those of mammography, suggesting that the biopsy rate of ultrasound after negative mammography is likely to be at least four times that of mammography alone. This is the major limitation of screening ultrasound. The PPV3, which represents the percentage of biopsies that are positive for cancer, was only 7% in studies of women with dense breasts and in the high risk population in the ACRIN 6666 study. The PPV3 in mammography is approximately 25%. Thus the rate of false positive biopsies is much higher with ultrasound. Table 11 below summarizes the key statistics from the three Connecticut studies (direct evidence) and the ACRIN 6666 study (high quality indirect evidence).

Table 11: Key findings from the essential, U.S.-based studies of HHUS.

Study	Recall Rate Per 1,000	Biopsy Rate Per 1,000	PPV3	Cancer Detection Rate Per 1,000
Hooley 2012	56.7	56.7	5.7%	3.2
Weigert 2012	49.6	48.3	6.7%	3.2
Parris 2013	33.5	32.8	5.5%	1.8
ACRIN 6666	185.7	88.0	6.8%	5.9

PPV3: positive predictive value of biopsies actually performed

The studies comparing mammography, ultrasound, and MRI in very high-risk women described in the section on MRI also help with the comparative effectiveness of the three technologies. In the six studies evaluating all three technologies (Kulh, 2005; Sardanelli, 2007; Warner, 2004; Berg, 2012; Lehman, 2005; Kuhl, 2010), mammography detected 48 cancers, ultrasound detected 53, and MRI detected 116. Ultrasound detected 19 cancers that were not detected by mammography, which represents 40%

(19/48) more cancers detected. Four cancers (3%) were detected only on ultrasound. These studies suggest that the addition of HHUS would increase the cancer detection rate by about 40% (best estimate: 1.4-fold increase) more than mammography alone, but that HHUS does not increase the cancer detection rate when added to mammography plus MRI.

There are no large, well-conducted studies in the United States that directly measure these statistics, which could serve as a reasonable estimate. There is also uncertainty about whether the early detection of these cancers by ultrasound will improve outcomes for women compared to outcomes following their detection as a lump by the women before her next mammogram (interval cancers) or when she has her next screening mammogram.

Estimates based on these data are shown in Table 12 below. There is a low level of uncertainty around the PPV3 because it was fairly consistent in the literature. The cancer detection rate comes primarily from the three studies describing the experience in Connecticut. There is high uncertainty about the recall rate because of the lack of direct evidence from studies of HHUS in women with dense breast tissue and because of the heterogeneity of the findings in the studies. In all of the prospective studies the recall rate was greater than 100 per 1,000 examinations.

Table 12: Estimated incremental yield of HHUS after negative digital mammography in women with dense breast tissue.

Statistic	Digital Mammography	Incremental Yield with HHUS	Uncertainty
Recall rate per 1,000	100-120	30-100	High
Biopsy rate per 1,000	14-22	30-60	Low-moderate
CDR per 1,000	3-5	2-4	Low
PPV3	20-25%	5-7%	Low

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

These estimates suggest that HHUS would find 2-4 more cancers than those found by digital mammography alone, with a PPV3 of 6-8%. Recall rates would range widely (30-100) and 30-60 additional biopsies would be required in order to identify these cancers.

8.5 Automated Whole Breast Ultrasound (ABUS)

Five studies (Kelly, 2010; Stoblen, 2011; Giuliano, 2013; Arleo; 2014; Brem, 2014) of ABUS evaluating approximately 9,000 participants met the inclusion criteria for the assessment. The primary data are summarized in Table 13 below; studies are described in detail in the sections that follow.

Table 13: Key findings from the available studies of ABUS.

Study	Recall Rate per 1,000	Biopsy Rate per 1,000	PPV3	Cancer Detection Rate per 1,000
Arleo 2014	188.2	19.7	0	0
Kelly 2010	75.8	12.2	NR	3.8
Brem 2014	150.2	36.0	9.8	7.3
Stoblen 2011	206.9	NR	NR	0
Giuliano 2013	22.8	NR	NR	12.3

PPV3: positive predictive value of biopsies actually performed

Kelly and colleagues (Kelly, 2010) recruited women from eight facilities across the United States. The investigators offered ABUS to consecutive asymptomatic women who had dense breasts. The radiologist reading the mammogram was blinded to the ABUS results and the radiologist reading the ABUS was blinded to the mammography results. Women whose compressed breast thickness at mammography was greater than 7 cm were excluded because of the limited sensitivity of ultrasound at that depth. The percentage of patients who agreed to participate at each site varied from 5% to 25%. The investigators performed 6,425 ABUS examinations in 4,419 women. 1,434 of the examinations were diagnostic examinations because the women had a history of prior breast cancer (776/1,434; 54%), breast implants (399/1,434; 28%), or non-localized abnormalities such as diffuse tenderness or nodularity (159/1,434; 11%). One third of the mammograms were digital and two-thirds were film. Some women at high risk preferred to alternate mammography and ABUS examinations at six-month intervals. The study followed women for one year for interval cancers. The percentage with complete follow-up was not reported, but 5,089 of the women (80%) had a repeat mammogram at least one year after the original mammogram.

The ABUS examination took five to 10 minutes preparation time and 10 to 20 minutes for the examination (Kelly, 2010). The interpretation and reporting time for the radiologist was seven to 10 minutes. The study sample had a median age of 53 years, but included women as young as age 24 and as old as 89 years. The sample included women with a personal history of breast cancer (10%), at least one first-degree relative with breast cancer (30%), and at least one second-degree relative with breast cancer (29%). These proportions are higher than in a typical screening population, suggesting that women who enrolled in the study were at higher risk for breast cancer than the general population.

During the study 23 breast cancers were detected with mammography, 23 by ABUS in women with negative mammograms, and an addition 11 presented as interval cancers that were not detected by either modality. One woman was diagnosed with bilateral breast cancer (56 participants diagnosed with

breast cancer). The results are not presented separately for women with negative mammograms, but the statistics for ABUS after a negative mammogram can be calculated from the data presented in the tables and the results section (see Appendix B). The sensitivity of mammography plus ABUS was higher than that of mammography alone (67.6% compared to 41.1%), but the recall rate for mammography plus ABUS was almost double that of mammography alone (74.8 per 1,000 compared to 32.4 per 1,000). The biopsy rate was also higher with mammography plus ABUS (12.2 compared to 9.1 per 1,000). The cancer detection rates were similar (3.8 compared to 3.6 per 1,000).

There are many methodological concerns that limit the ability to generalize the results of this study to women with dense breasts and a negative digital mammogram. The low volunteer rate (5%-25%) in this study raises concerns about spectrum bias – those who agree to participate may differ from those who do not participate in ways that impact the study results. For instance, women at higher risk for breast cancer may be more likely to volunteer for this study of additional imaging. The wide age range (down to age 27 years) also suggests that this was not a typical screening population. Two-thirds of the mammograms were film, which has much lower sensitivity than digital mammography in women with dense breasts – some of the cancers identified on ABUS would have been picked up by digital mammography (Kerlikowske, 2011; Pisano, 2005). In addition, some of the women elected to be screened with ABUS six months after the mammogram, so the cancers identified by ABUS may represent a mix of interval cancers and those missed by mammography. All of these biases would tend to increase the cancer yield of ABUS.

In the second, much smaller study (Stoblen, 2011), Stoblen and colleagues described the results of ABUS in 304 consecutive women between the ages of 50 and 69 who were seen for routine screening mammography in Germany. The majority of the women had non-dense breasts (scattered fibroglandular densities). All subjects had digital mammography followed by ABUS. Two cases of DCIS were detected by mammography, neither of which was detected by ultrasound. The investigators reported 60 false-positive assessments by ABUS (20.7% of negative mammograms) compared to 12 (4.0%) for digital mammography in the same women. Thus the false positive rate for ABUS was 207 per 1,000 examinations compared to 40 per 1,000 for mammography. However, it does not appear that all of the positive ultrasound findings were biopsied. In addition, no follow-up was reported other than for two patients with repeat examinations at six months, with no additional cancers identified. The study is small and does not directly apply to women with dense breasts, but it highlights the concern about high numbers of false positive results with either automated or hand held ultrasound.

Giuliano and Giuliano (2013) report on the performance of digital mammography plus ABUS in 3,418 asymptomatic U.S. women with dense breasts, compared to 4,076 asymptomatic women with dense breasts screened with digital mammography in the prior year. It is unclear if consecutive women were included. The BI-RADS categories were not used to define high density, but it is likely that the women studied (mammograms with “a Wolfe classification of 50% or greater”) were similar to the two high-density BI-RADS groups. The study excluded women with major risk factors for breast cancer including those with a personal or family history of breast cancer and those with a BRCA mutation. The study was performed at a single site in Florida. Two radiologists read each of mammograms and the ABUS images with final readings by consensus. There was no blinding of the radiologists, but the investigators blinded the pathologists evaluating biopsy specimens.

In the control group, the sensitivity and specificity of digital mammography alone were 76.0% and 98.2% (Giuliano, 2013). The recall rate was 22.8 per 1,000 examinations and the cancer detection rate was 4.7 per 1,000. The biopsy rate was not reported. This cancer detection rate is relatively high for invasive cancer (no cases of DCIS were reported) in women with no personal or family history of breast cancer.

The PPV reported (20.4%) is quite high for mammography, suggesting that it may be the PPV for biopsy rather than the PPV for a positive mammography assessment. The low recall rate also supports the under-reporting of recalls for positive results. In the mammography plus ABUS group, the sensitivity was 97.7% and the specificity was 99.7%. Again, the recall rate and biopsy rate are not clearly reported and both are calculated at 15.2 per 1,000 examinations, suggesting that this is actually the biopsy rate. It is likely that the true recall rate was much higher and that the specificity and PPV are much lower than reported in the paper.

There are several other concerns about this study that call into question all of its results. First, there were no reported cases of DCIS. In 2011, when the study was conducted, approximately 27% of all breast cancer diagnoses were DCIS (DeSantis, 2011). Since the study reported 68 invasive breast cancers, there should have been an additional 25 cases of DCIS. Mammography is more sensitive than ultrasound for the detection of DCIS (Kelly, 2011), so the exclusion of DCIS from the results could have a large impact on the results. It is also worrisome that the results for mammography alone were not reported for the cohort of women also examined with ABUS. It may be that digital mammography performed better in that group because the radiologists had one more year of experience with this relatively new technology. It is also remarkable that the specificity of ABUS was so high. All other reports of ultrasound consistently find a high rate of false positive studies with ultrasound with PPV1 and PPV3 being consistently lower than that of mammography. The opposite was reported in this study.

The demographic characteristics of the two groups were not presented, nor compared – if they were very different, then there should have been some adjustment for these differences. The results suggest that there were large differences: the average age for detected invasive cancers in the control group was 54 years, while in the ABUS group it was 57. If ABUS identifies cancers earlier than mammography, then the average age of detected cancers should go down, not up. Given these major concerns, as well as the non-standard breast density measurements and the lack of reporting of the results of ABUS among the women with normal mammograms, the results of this study are not particularly useful in evaluating the appropriate role for ABUS in women with dense breasts.

A small, single-center retrospective study (Arleo, 2013) evaluated 558 ABUS examinations over three months (August-October 2013). The aim of the study was to evaluate differences in the recall rate after implementation of an ABUS system in women with dense breasts (BI-RADS category 3 or 4); if women had been screened previously at this center, results were required to be normal or benign. Women presenting with symptoms or with prior positive ultrasound were screened with HHUS instead. Of the 558 examinations performed on 558 women, 10 women had a family history, three had implants, and 15 had a history of benign breast biopsies. A total of 105 women were recalled for additional screening with HHUS (18.8 recalls per 1,000). The authors noted that this rate improved substantially over the course of three months – from 24.7 per 1,000 in August to 12.6 per 1,000 in October – suggesting an initial learning curve for the new technology, and potential to improve with experience. Eleven women were biopsied, all with benign findings (i.e., PPV3 of 0%). Because of the study's short duration, no other performance characteristics were calculated.

There are some methodological issues with this study beyond the small sample size and short duration. First, because the study evaluated women during the initial implementation of ABUS, the capability of ABUS to reduce the recall rate cannot be fully understood from this analysis alone. Perhaps more importantly, however, there are no data presented on how many women had a prior normal or benign screening result, and how much time had elapsed between that screen and ABUS. It is therefore impossible to put the recall rate presented here in its proper context.

Finally, a very recent prospective multinational study (Brem, 2014) evaluated the use of ABUS plus mammography compared to digital mammography alone in 15,318 women and found that adjunctive ABUS yielded a cancer detection rate of 7.3 cancers per 1,000 women screened compared with 5.4 cancers for screening mammography alone (p<.001). However, the recall rate increased from 150.2 recalls per 1,000 women screened with mammography to 284.9 recalls for the combined approach (p<.001). PPV1 decreased when adding ABUS (2.6% vs. 3.6% with mammography alone), and required an additional 36 biopsies per 1,000 women screened. Sensitivity improved significantly by adding ABUS (100% vs. 73.2% for mammography alone), but specificity was reduced (72.0% vs. 85.4%) (p<.001 for both comparisons). Although this study included adequate follow-up and reported results for nearly all outcomes of interest, the initial mammography screenings were not performed by radiologists, which may have introduced some bias by over-assigning women into dense breast categories on the subsequent reading.

Summary: Screening ABUS

None of the studies directly address the use of ABUS following negative digital mammography in a screening population of women with dense breasts. Four of the studies are of poor quality, and only the recent Brem study is of fair quality. Some of these studies (e.g., Kelly, 2010; Brem, 2014) offer reasonable estimates and relatively complete reporting of outcomes, but are of limited applicability because of technical concerns (e.g., use of film mammography, long gaps between mammography and ABUS, use of research coordinators and technologists to assign breast density categories). Across the five studies, the recall rate varied from 5 to 285 per 1,000 examinations, the biopsy rate was not reported or up to 15 per 1,000 examinations, the PPV3 from not reported to 31% and the cancer detection rate ranged from 0 to 7.6 per 1,000 examinations. Overall, the paucity of studies, the lack of high quality studies, and the wide range of estimates across the studies mean that there is considerable uncertainty surrounding all of the estimates for the diagnostic test statistics for ABUS.

Because of the uncertainty described above, we felt that the most reliable estimates for the test characteristics for ABUS come from the HHUS literature, but with high uncertainty. Estimates based on these data are shown in Table 14 below and are identical to those for HHUS.

Table 14: Estimated incremental yield of ABUS after negative digital mammography in women with dense breast tissue.

Statistic	Digital Mammography	Incremental Yield with ABUS	Uncertainty
Recall rate per 1,000	100-120	30-100	High
Biopsy rate per 1,000	14-22	30-60	High
CDR per 1,000	3-5	2-4	High
PPV3	20-25%	5-7%	High

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

These estimates suggest that ABUS would find 2-4 more cancers than those found by digital mammography alone with a PPV3 of 5-7%. Recall rates would range widely (30-100) and 30-60 additional biopsies would be required in order to identify these cancers.

Key Question #3: What are the documented and potential harms associated with these imaging tests, including overdiagnosis and overtreatment, unnecessary biopsy as a result of false-positive imaging, patient anxiety, and radiation exposure?

8.6 Potential Harms

Overdiagnosis and Overtreatment

An important harm of screening is overdiagnosis: the diagnosis of breast cancers with mammography that, if they had been left undetected, would not have caused symptoms before the woman died of other causes (Zahl, 2008; Welch, 2010). Such patients would endure the toxicity associated with overtreatment of breast cancer (surgery, radiation, hormonal therapy, and chemotherapy), without receiving any benefit of reduced symptoms or longer life from treating the cancer.

None of the studies in our review attempted to measure rates of overdiagnosis or overtreatment based on test performance; for that matter, no long-term clinical outcomes of interest were tracked. This is largely because it is currently impossible to know whether any particular patient whose cancer is detected by mammography is or is not at risk of the cancer being overdiagnosed, and the true magnitude of "overdiagnosis" and subsequent "overtreatment" for breast cancer is unclear and controversial. The most common estimates range from 10% to 30% of cancer diagnoses, although estimates range from as low as 0% to as high as 54% (Blyer, 2005; Welch, 2009a; Jorgensen, 2009; Welch, 2009b; Morrell, 2010; de Gelder, 2011; Wu, 2011; Elmore, 2012; Kalager, 2012; Gotzsche, 2013). This is an area of active research and debate.

Unnecessary Biopsy

The most common harm associated with mammography is a false-positive test result. Approximately 10% of women have a false-positive result at each round of mammography screening and about 50% of women will have at least one false-positive result after 10 mammograms (Rosenberg, 2006; Chrisiansen, 2000; Elmore, 1998; Hofvind, 2004; Olivotto, 1998; Hubbard, 2011). False positives also usually require that a woman schedule a second appointment for additional imaging resulting in time lost with family or at work and the additional evaluation increases health care costs. Most false-positive results do not lead to a breast biopsy; between 7% and 19% of women have a false positive biopsy after 10 mammograms (Elmore, 1998; Hofvind, 2004).

With regard to the tests of interest in this analysis, biopsy rates ranged from a low of 12 per 1,000 for DBT to up to 60 per 1,000 for ultrasound. Because the overall risk of breast cancer in any large screening population is relatively low, most of these will be unnecessary (i.e., false-positives); for supplemental tests, this will only add to the burden of false-positives already produced by mammography. While false-positive biopsies have effects on patient anxiety (see below), they are also not without clinical consequence. A recent AHRQ review (Dahabreh, 2014) found that severe complications were rare for all forms of breast biopsy (<1%), but that harms were also generally underreported across all studies. Complications can include local reactions, bleeding, and infection or abscess; there have also been case reports of tumor formation at biopsy sites.

Patient Anxiety

While not well-documented in studies of imaging tests, many patients experience short-term increases in anxiety and psychological stress as a result of being recalled after an abnormal finding (Barton, 2001; Barton, 2004; Lipkus, 2000; Scaf-Klomp, 1997; Brett, 2005; Weil, 1997; Woodward, 2001). A systematic review of 23 studies on the long-term effects of false positive mammograms found small, but significant negative impacts on health behaviors and psychological well-being (Brewer, 2007). A recently published case series (Miller, 2013) evaluated women undergoing image-guided breast biopsy and found that stronger physician-patient communication reduced levels of patient anxiety and improved health-related quality of life. Recent advances in online support tools have also shown that increasing interaction between patients and clinicians eases anxiety in women who experience an abnormal screening (Obadina, 2014). Since no imaging modality can be 100% accurate, it is important to consider the use of these anxiety-reducing interventions when weighing the harms and benefits of screening tests.

Radiation Exposure

Ionizing radiation, like that used in mammography or DBT, can damage DNA leading to mutations that increase the risk for the development of cancer. Evidence from those exposed to radiation from the atomic bomb explosions in Japan and from those exposed to radiation therapy as part of treatment for Hodgkin's disease demonstrates that radiation exposure increases the risk for breast cancer (Boice, 2001, Carmichael, 2003; de Gelder, 2011; Ng, 2009; Ronckers, 2005; Yaffe, 2011). The risk is greatest for younger women and is thought to be minimal for post-menopausal women. The radiation dose from mammography is relatively small. The dose from 20 mammograms is equivalent to about three years of environmental exposure to radiation; the dose from one CT scan is equivalent to about 800 mammograms. There is no direct evidence demonstrating an increase in breast cancer due to mammography. One recent modeling study by Yaffe and colleagues estimated that among 100,000 women screened with mammography every year from ages 40 to 55 years and then every two years until age 75 (20 mammograms), the radiation would cause 86 new breast cancer diagnoses and 11 deaths from breast cancer (Yaffe, 2011). Thus for every 1,000 women screened 20 times between the ages of 40 and 75 years, the radiation from mammography may cause 0.9 additional breast cancers and 0.1 additional deaths from breast cancer.

The average dose of radiation from mammography has declined with the transition to digital mammography. In the DMIST trial, the average radiation dose was 4.7 mGy with film mammography and 3.7 mGy with digital mammography (Hendrick, 2010). The Yaffe model assumed that the dose per mammogram was 3.7 mGy based on the DMIST findings (Yaffe, 2011; Hendrick, 2010). Other models using different inputs and assumptions have estimated higher rates of radiation-induced breast cancer and death from mammography (Berrington, 2005).

Although some concerns about the additional exposure to ionizing radiation (double the amount of digital mammography alone) with DBT were initially raised, recent developments in computer software designed to work with existing DBT systems to produce synthesized 2D images from advanced 3D acquisition data have abrogated the need for screening with digital mammography and reduced the radiation exposure to the patient by 50% (Zuley, 2014; Skaane, 2014). The FDA approved the use of such software in May 2013 so evaluation its impact on cancer-related outcomes has not yet been widely studied; nevertheless, once this approach is standardized, it is likely that radiation exposure from DBT will be comparable to that of digital mammography.

Key Question #4: What is the differential effectiveness and safety of the tests of interest according to such factors as age, race or ethnicity, comorbidities, BMI, method of breast density classification, overall breast cancer risk, scan vendor, and imaging protocol (e.g., whether ultrasound is performed by a radiologist, technologist, or some combination of the two)?

8.7 Differential Test Performance in Key Subgroups

Limited data exist to support the use of imaging tests in specific subpopulations beyond the ones included in this assessment (i.e., women with dense breasts or high risk populations). The most common subgroup analyzed in the literature on DBT was age. Three studies (Rose, 2013; Haas, 2013; Destounis, 2014) included subgroup analyses on breast density and age, and both reported similar improvements with DBT for the evaluated outcomes across age groups. One of these studies (Rose, 2013) also evaluated a subgroup of women undergoing screening for the first time, as well as the impact of radiologist experience; however, no material differences in study outcomes were observed. Although no DBT studies evaluated performance across multiple vendors, an editorial observed that the most recent study included in this assessment (Friedewald, 2014) did not use the newest DBT technology, and so considerations of its performance are already outdated (Pisano, 2014).

In the studies evaluating automated whole-breast ultrasound, differential effectiveness related to technologist experience was assessed. Because ultrasonography requires a high degree of technical and medical proficiency, analyzing the imaging protocol of this procedure is important in these early observations (Stoblen, 2011). In a single-center retrospective study (Arleo, 2013) evaluating the first calendar quarter after implementation of ABUS, recall rate improved from 24.7% in the first month to 12.6% at the end of the third. Because ABUS can be operated independently, the investigators suggested an advantage with ABUS over HHUS, particularly with additional experience over time (Arleo, 2013). Moreover, because the automated procedure eliminates operator variability, whole-breast ultrasonography may be more easily reproducible and operators are more confident in their ability to recommend further screening (Kelly, 2010). As noted in the body of the evidence review, however, currently-available ABUS data are too limited to make even indirect comparisons to the other modalities of interest.

One additional study, a retrospective examination of HHUS in Korea (Chae, 2013) examined the impact of age on test performance. Sensitivity was comparable for mammography+HHUS versus mammography alone in all age groups except women age 40-49 (100% versus 29%, $p<.05$). It is difficult to generalize these results to U.S. settings, however, as (a) the overall incidence of breast cancer in Asian countries is less than that in the U.S.; and (b) in contrast to the U.S., where breast cancer incidence increases with age, incidence in Korea is highest among women in their 40s (Chae, 2013).

8.8 Summary

Mammography is the only screening test that has been shown to reduce breast cancer mortality in randomized trials (Alexander, 1997; Andersson, 1988; Andersson, 1997; Bjurstam, 1997a, Bjurstam, 1997b; Frisell, 1997; Miller, 2000; Miller, 2002; Nystrom, 2002; Shapiro, 1988; Tabar, 2000; Moss, 2006). However, it is not perfect. At best, the sensitivity of mammography, including digital mammography, is

approximately 80% (Rosenber, 2006; Pisano, 2005; Kerlikowske, 2011). Thus for every four to five breast cancers detected on mammography, an additional interval breast cancer will be diagnosed prior to the next screening mammogram. Furthermore, to diagnose those cancers, many women will be recalled for additional imaging because of false positive assessments, and some of those women will undergo breast biopsy. Using current digital mammographic techniques in the United States, it can be estimated that for every 1,000 women having a screening mammogram, approximately 100 will be recalled for additional tests, 10 will have a breast biopsy, five will be diagnosed with breast cancer, and one additional cancer will be diagnosed in the subsequent year. The false positive mammography results lead to additional time lost for the women who must schedule time to come in for additional tests and adds cost to the medical system. The women may also experience unnecessary anxiety about a cancer diagnosis.

Radiologists have long known that areas of density in the breast can obscure breast cancers on film mammography leading to a false negative assessment (decreased sensitivity). Across the four categories of breast density, the sensitivity of film mammography decreases from about 85% for women in the two lowest density categories to approximately 80% for women with heterogeneously dense breast tissue, and 65% for women with extremely dense breasts (Carney, 2003; Kerlikowske, 2011). This masking effect of breast density is one of the primary reasons that state legislatures have passed laws requiring that women be notified about their breast density if they are in one of the high density categories.

Over the past decade, however, film mammography has been replaced by digital mammography. Digital mammography has a higher dynamic range than film and greater contrast resolution allowing the display of more gradations of density when a radiologist views the image on a computer screen. One of the strengths of digital mammography is improved sensitivity for breast cancer in dense breast tissue. In the DMIST trial, which assessed women with both film and digital mammography, the sensitivity of mammography in the two high-density categories was 55% for film mammography but 70% for digital (Pisano, 2005).

In the BCSC, a large registry of woman screened for breast cancer, the sensitivity of digital mammography was approximately 80% to 85% across all four breast density categories, with no trend towards a decrease in sensitivity with increasing breast density (Carney, 2003). Thus, the risk of masking has been dramatically reduced by the widespread adoption of digital mammography. Nonetheless, even without masking, approximately one in five cancers can still be missed by digital mammography, raising questions about the potential for benefits of additional screening, especially among women at highest risk for breast cancer.

It appears that another shift is on the horizon—from digital mammography to DBT. This technology decreased the recall rate in the six comparative studies considered in this assessment (Ciatto, 2013; Skaane, 2013a; Skaane, 2013b; Haas, 2013; Rose, 2013; Friedewald, 2014; Destounis, 2014). At the same time, DBT increased the cancer detection rate by about 1-2 per 1,000 examinations compared to digital mammography alone. One of the studies also reported that the biopsy rate decreased from 15.2 to 10.6 per 1,000 examinations (Rose, 2013), although the largest retrospective study conducted to date reported an *increase* in the biopsy rate from 18.1 to 19.3 per 1,000 (Friedewald, 2014). In the subgroup of women with dense breasts, DBT identified an additional 2.7 cancers per 1,000 examinations with a recall rate of 21.3 per 1,000 examinations. DBT has the advantage of being easy to incorporate into the process for routine mammography screening, requiring little extra time from the woman being screened (Skaane, 2013a), and recent technological advances suggest that this can be done without excess radiation (Houssami, 2013; Skaane, 2013a). There are also technical aspects that are still under

development, such as accurate biopsy techniques for abnormalities identified on DBT, but not visible on the digital mammogram (Viala, 2013).

More importantly, however, the literature on DBT remains incomplete with regard to true population sensitivity, specificity, and negative predictive value, as only one of the six available studies had sufficient follow-up to assess interval cancers, and that study had a 20% dropout rate before one year (Destounis, 2014). Still, it appears that DBT holds the promise of reduced recalls with comparable or better cancer detection than digital mammography alone.

The available literature has shown that all three of the advanced imaging technologies evaluated in this assessment can detect additional breast cancers in women with negative mammograms. The most convincing data on cancer detection rates come from the ACRIN 6666 trial (Berg, 2012): in the third round of screening, the combination of mammography and HHUS detected seven additional cancers (cancer detection rate 11.4 per 1,000), and MRI detected an additional nine cancers (incremental cancer detection rate 18.2 per 1,000). However, the addition of MRI increased the number or recalls from 100 to 159 and the number of recommended biopsies from 38 to 81. The PPV3 of MRI in women with negative mammograms (22%) was much higher than that of HHUS in women with negative mammograms (7%). Thus the yield per biopsy of MRI was higher than HHUS. If the costs and logistics of the two were identical, MRI would be preferred as it has greater cancer detection with fewer harms from false positive biopsies. These results are similar to earlier studies that compared mammography, HHUS, and MRI in women at very high risk for breast cancer (Kuhl, 2005; Sardanelli, 2007; Warner, 2004; Lehman, 2007; Kuhl, 2010). However, there is little direct evidence about the utility of MRI in the population that is the focus of this assessment: women with dense breasts and a negative mammography assessment. MRI also requires an IV, carries the risk of complications from the injection of the contrast agent, and is the most time-consuming and expensive option.

The incremental cancer detection rate is about three per 1,000 examinations for HHUS versus mammography alone. However, HHUS markedly increases the recall rate and biopsy rate. There are far more studies on HHUS than the other technologies, but the study results vary dramatically, which introduces considerable uncertainty into the estimates of the potential impact of supplementary HHUS for women with dense breast tissue. HHUS approximately doubles the recall rate of mammography alone and quadruples the biopsy rate. HHUS has the advantage of being readily available at most breast imaging centers and does not utilize ionizing radiation. However, HHUS requires substantial training and experience of the technicians and radiologists to guarantee high quality results and it involves a substantial investment in radiologists' time.

Finally, there are extremely limited data on screening ABUS (Kelly, 2010; Stoblen, 2011; Giuliano, 2013; Arleo, 2014; Brem, 2014). The incremental cancer detection rates ranged widely, from 0 to 7.6 per 1,000 examinations. One of the studies reported a reduction in the recall rate with ABUS (Giuliano, 2013), but the other two had substantial recall rates that were equivalent to those seen with HHUS (Kelly, 2010; Stoblen, 2012). Two of the studies (Kelly, 2010; Giuliano, 2013) had a low biopsy rate and a high PPV3 suggesting that very few women are inappropriately being referred for biopsy. ABUS also has the advantage of little operator dependency, which addresses one of the major concerns with HHUS (Arleo, 2014).

Table 15 on the following page summarizes the estimates for each of the technologies of interest based on the clinical data published through August 2014. Note that mammography and DBT statistics are for all women, while HHUS/ABUS and MRI are for the subset of women with dense breast tissue and a negative mammogram. Many of the estimates have a high degree of uncertainty and will likely change

as more high quality data become available. However, they provide reasonable estimates of the clinical benefits and harms relative to each other

Table 15: Summary of the key statistics for digital mammography, digital breast tomosynthesis, and three supplemental screening technologies in women with dense breast tissue.

Statistic	DM	DBT	HHUS/ABUS*	MRI*
Recall rate	100-120	80-90	30-100	100-120
Biopsy rate	14-22	12-27	30-60	20-40
CDR	3-5	4-6	2-4	3-11
PPV3	20-25%	25-30%	5-7%	22-48%

*In the subset of women with dense breast tissue and a negative mammogram

CDR: cancer detection rate; PPV3: positive predictive value of biopsies actually performed

Table 15 also highlights the low PPV3 of ultrasound compared to the other technologies, which translates into a large number of unnecessary biopsies for every cancer detected by ultrasound. Table 15 also clearly illustrates that DBT has a lower recall rate than digital mammography and that MRI detects the greatest number of additional breast cancers.

Thus, we know with a high degree of certainty that both DBT and all forms of supplemental screening find additional breast cancers. Most of the cancers are small, lymph node negative, and thus are potentially curable. MRI finds the most cancers, DBT has the lowest recall rate, and HHUS results in the largest number of false positive biopsies.

The major unanswered question is whether the identification of additional cancers through DBT and/or supplemental screening improves outcomes for women. Some advocates of supplemental screening will argue that the majority of the cancers identified through supplemental screening are early stage cancers with an excellent prognosis following treatment. These represent the spectrum of cancers identified with mammography that led to the reduction in mortality seen with the randomized trials of screening mammography. In their view, there can be no question that patient outcomes will be improved with supplemental screening. Others will argue that many of these supplemental screen-detected cancers would have been cured when detected on physical exam or subsequent screening mammograms and that some of these cancers represent overdiagnosis, which leads to net harm for the patient. They will highlight the growing evidence for significant overdiagnosis with mammography alone. These individuals will suggest that much of the incremental cancer detection rate with HHUS (2 to 4 or more per 1,000 examinations), which is much higher than the expected interval cancer rate (about 1 per 1,000 examinations), can only represent overdiagnosis. Only large, long-term randomized trials can definitively answer this question.

Key Question #5: What are the costs and cost effectiveness (e.g., cost per cancer detected) of the imaging modalities of interest?

9. Model of Clinical and Economic Outcomes of Breast Cancer Screening

As noted in this review, published evidence on the clinical effects of (a) digital breast tomosynthesis for general population breast cancer screening; and (b) supplemental screening in women with dense breast tissue is quite limited; in fact, the target population of interest for the latter aim (women with BI-RADS "c" or "d" breast density and a negative mammogram) has been used extensively only to evaluate HHUS. We developed a cohort model to address these gaps, focusing on the clinical and economic outcomes both for replacing mammography with tomosynthesis for the general population and supplemental screening among eligible women in the state of Washington. Supplemental screening modalities considered included HHUS, ABUS, and MRI, used among women with dense breast tissue and a negative mammogram, or alternatively, a negative DBT.

Information on the economic impact of DBT or of any screening strategy for women with dense breast tissue is also quite limited. We nevertheless summarize the published evidence relevant to the scope of this review in the section below.

9.1 Prior Published Evidence on Costs and Cost-Effectiveness

DBT vs. Digital Mammography

A single, recently-published study examined the cost-effectiveness of screening with DBT vs. digital mammography (Lee, 2014). A discrete-event breast cancer simulation model developed as part of the National Cancer Institute's Cancer Intervention and Surveillance Modeling Network (CISNET) was used to examine the outcomes and costs of biennial screening with the combination of DBT and digital mammography vs. digital mammography alone among women aged 50-74 years with heterogeneously or extremely dense breast tissue. The test performance of DBT was assumed to be moderately improved over digital mammography (sensitivity 80%, specificity 92% vs. 77%/88% for mammography). DBT screening costs were estimated to be $50 higher for DBT than for digital mammography alone. On a lifetime basis, the use of DBT resulted in 0.5 fewer deaths and 405 fewer false-positives per 1,000 women screened after 12 rounds of screening. Discounted costs were approximately $350 higher for DBT, which delivered an incremental 0.007 QALYs (approximately 3 days) and a cost-effectiveness ratio of approximately $54,000 per QALY gained. Findings were relatively robust to changes in estimated test performance and the disutility of false-positives as well as the additional cost of DBT—cost-effectiveness remained below $100,000 per QALY gained at incremental costs of up to $87 for DBT.

In addition, other studies have examined different strategies for screening mammography in women with dense breast tissue and are summarized here for additional context. Tosteson and colleagues used a simulation model to evaluate the cost-effectiveness of different screening strategies using digital mammography vs. film mammography for U.S. women aged 40 and older, using data from the DMIST trial and other sources (Tosteson, 2008). A strategy of targeted digital mammography (i.e., either for women age <50 and/or women of any age with dense breasts, film in all other women) was estimated to produce more screen-detected cases of cancer and fewer cancer-related deaths than either an all-film or all-digital strategy. Estimates of cost-effectiveness were $26,500 per quality-adjusted life year (QALY) gained for age-targeted digital mammography vs. all-film mammography and $84,500 per QALY for age- and density-targeted digital vs. all-film. A density-targeted digital strategy focused on the Medicare population (age ≥65) yielded cost-effectiveness estimates ranging from $97,000 - $257,000 per QALY gained vs. all-film, depending on assumptions regarding the test performance of digital vs. film mammography.

A more recent study evaluated the performance of multiple digital mammography screening strategies in a U.S. cohort, based on findings from four distinct simulation models (Stout, 2014). Extending annual screening to women age 40-74 with dense breast tissue (with those with non-dense tissue screened biennially) resulted in cost-effectiveness ratios ranging from ~$60,000 - $260,000 per QALY gained relative to screening all women age 40-74 biennially. Variability in model findings was attributed to relatively small incremental benefits from each screening strategy, coupled with high model sensitivity to assumptions regarding women's preferences for avoiding false-positive results. Consistent with the Tosteson study, annual screening strategies in these models that were not targeted by age and/or breast density were not found to be cost-effective.

However, a third study employed a Markov model to compare biennial and annual screening mammography among women with dense breast tissue in Canada (Pataky, 2014), and found that, compared with a biennial approach, annual screening produced a very small increase in QALYs (0.0014, or less than one day) with increased costs of over $800 per patient. An annual approach produced a cost-effectiveness ratio of approximately $570,000 CAD ($510,000) per QALY gained vs. biennial screening, and had a 37.5% probability of being cost-effective at a willingness-to-pay threshold of $100,000 per QALY gained.

Supplemental Screening in Women with Dense Breast Tissue

Economic evaluations of supplemental screening strategies have directly assessed costs using primary data collected from cohorts of women undergoing screening. One of these reported actual cancer detection and costs from a series of 5,227 asymptomatic Italian women with dense breast tissue and negative mammograms who had HHUS within one month of film mammography.[149] Costs included those of HHUS, clinical examination, biopsy, and cytology, and totaled €56 ($77) per HHUS-screened woman. HHUS detected two additional cancers in this cohort (0.4 per 1,000), resulting in a cost per additional cancer detected estimate of €146,497 ($200,701). The authors hypothesize that the cancer detection rate observed in this study, which was much lower than that reported in the HHUS studies summarized in this review (range: 1.8 – 14.2 per 1,000), may have been a result of self-selection. The sample was limited to women who presented for HHUS within one month of negative mammography, which represented approximately 20% of all women screened at the study site who had dense breasts and negative mammograms. In addition, 72% of women in the study sample were age <50, which is not reflective of the age distribution of women in the general screening population or of the subset with dense breast tissue.

Data are also available from two of the three cohort studies reporting ultrasound experience following the passage of Connecticut's breast density legislation (Hooley, 2012; Weigert, 2012). Hooley and colleagues estimated the cost of providing breast ultrasound to a cohort of 935 mammographically-negative women with dense breast tissue who were screened with HHUS at Yale-New Haven Hospital after passage of the law (Hooley, 2012; Tosteson, 2008). The incremental cancer detection rate was 3.2 per 1,000 screened. Costs, including those of HHUS, aspiration, and biopsy, totaled approximately $180,000 for the cohort, or $60,000 per additional case of cancer detected.

A larger retrospective study of HHUS screening in nearly 9,000 women with dense breast tissue and negative screened mammograms was conducted in 6 radiology practices in Connecticut for the year after passage of the breast density legislation (Weigert, 2012). The cancer detection rate was also 3.2 per 1,000 screened in this study. Costs were estimated for screening and biopsy based on billed charges

to insurers, and totaled approximately $3.1 million, or $110,000 per additional cancer detected. Neither study compared screening costs after passage of the law to costs incurred before the law was passed.

9.2 Overview of the Cohort Model

As described above, the published literature on the clinical and economic impact of DBT in any population, and of supplemental breast cancer screening modalities in women with dense breast tissue, is noticeably limited. We therefore developed a cohort model to perform a population-based, one-year analysis of clinical and economic outcomes specific to the state of Washington. In the model we included all women age 40-74 except for those with certain high-risk factors, including genetic susceptibility, personal history of breast cancer, and prior chest radiation (see "Target Population" below). Outcomes and costs included those of screening, diagnostic workup (including biopsy when performed), and detection and workup of interval cancers. Costs of treatment and other measures beyond one year of follow-up were not considered.

As this review has highlighted, the performance of digital mammography varies according to level of breast density. We first conducted baseline analyses comparing the one-year screening performance and costs for both digital mammography and DBT for all women undergoing screening. Because available DBT studies are both incomplete with respect to measurement of sensitivity and specificity and lacking detail on DBT's performance by category of breast density, we tested various possible levels of improvement in test performance relative to digital mammography in our analyses. In contrast to Lee's evaluation described above, we focused attention on the use of DBT as a frontline screening tool for *all* women (not only those with dense breasts), and the implementation of screening on an annual rather than biennial basis. While there have been recent studies suggesting that biennial screening may be more appropriate for women of all ages (Pataky, 2014; O'Donoghue, 2014), observational study suggests that over 60% of women who currently present for screening receive mammography annually (Kerlikowske, 2013).

Then, we used the model to compare the performance and costs of supplemental screening with each of the modalities of interest (i.e., HHUS, ABUS, and MRI) for women in BI-RADS density categories "c" or "d" who had an initial negative mammogram. For these analyses of supplemental screening, digital mammography was assumed for initial screening, as evidence indicates it is the current screening standard. However, we also examined the performance of these supplemental modalities in which DBT was the screening standard, and women with a negative DBT would then go on for further screening. As with frontline screening, outcomes and costs of supplemental screening were tracked over one year.

We defined the supplemental screening population as a hypothetical cohort that was stratified into different levels of underlying breast cancer risk. Specifically, we divided risk into three levels (low, moderate, and high) that would be based on the woman's age, breast density, and family history of breast cancer -- information likely to be available through physician-patient discussion in the primary care setting. Several more sophisticated risk assessment algorithms are available, but for modeling purposes we opted to use a simplified risk algorithm based on just these three factors to maximize the feasibility and potential generalizability of this approach (see "Overall Breast Cancer Risk" below).

We had to make several broad assumptions in designing the model that are important because they limit the ability of the model to capture the nuances of patient behavior and the many variations in clinical care patterns that occur for individual patients. For example, we assumed perfect compliance

for both mammography/DBT and supplemental screening in this analysis. While it is the case that actual compliance is always less than 100%, differences across studies in the definition of the time interval within which women are considered compliant as well as considerations of what constitutes screening vs. diagnostic mammography (Partin, 1998) precluded our use of a uniform, widely-accepted estimate for compliance across different imaging modalities.

The model also assumes that supplemental screening would occur immediately after a negative mammography or DBT result, and that one year of follow-up is available as the reference standard for both mammography/DBT and supplemental screening results. For mammography and DBT, we needed to estimate as "inputs" several important numbers based on our review of the clinical evidence, including the number of cancers detected (i.e., true positives), cancers missed (i.e., interval cancers), recalls for further testing, biopsies performed, cancer "yield" per biopsy (i.e., percentage of biopsies with positive results) and false-positive results both after biopsy and without biopsy (i.e., recalled for further testing but no biopsy recommended). We developed similar inputs for each supplemental screening strategy, but we made a simplifying assumption that all positive supplemental screening tests would result in immediate biopsy, and so did not estimate recall rates (which would equal biopsy rates in this case) or false-positive results without biopsy. As noted in this review, supplemental screening has the potential to detect both cancers missed by mammography or DBT and additional cancers that would not have presented during the interval between mammography/DBT screenings; we therefore included both types of cancer in our estimates for each supplemental modality.

Target Population

The population we modeled included all women age 40-74 except for those with known genetic susceptibility, a personal history of breast cancer, and/or a history of mantle radiation to the chest. We calculated the initial screening population size based on age- and gender-specific US Census data for Washington (U.S. Census Bureau, 2014). The prevalence of the risk factors noted above was estimated to total 4.7% in the general screening population, and we reduced the population size accordingly (see Table 16 on the following page for population-based model parameters and data sources). The resulting target population size was approximately 1.3 million Washington women. The distribution of BI-RADS breast density within each age band was estimated based on data from a recent BCSC publication (Kerlikowske, 2013). On a population basis, approximately 47% of the Washington screening population would have heterogeneously dense (~500,000) or extremely dense (~115,000) breast tissue.

Overall Breast Cancer Risk

As described above, we limited risk factors for breast cancer in our assignment of risk category to age, breast density, and close family history (at least one 1st degree relative), factors that would be readily available for discussion in a primary care setting. The percentage of women with dense breast tissue *and* a close family history was estimated to be 22.7% based on data from a New Hampshire mammography registry study (Titus-Ernstoff, 2006). Using these three risk factors alone in the BCSC Risk Calculator (Tice, 2008), we defined categories of low, moderate, and high risk as below:

Low: BI-RADS density 3 or 4, age 40-49, no close family history (corresponds to 5-year risks generally <1.7%). Risk assumed in the model: 1% (0.2% per year)

Moderate: BI-RADS density 3 or 4, age 40-49, with a close family history; OR BI-RADS density 3 or 4, age 50-74, no close family history (corresponds to 5-year risks generally between 1.7% and 3.0%). Risk assumed in the model: 2.5% (0.5% per year)

High: BI-RADS density 3 or 4, age 50-74, with a close family history (corresponds to 5-year risks generally >3.0%). Risk assumed in the model: 5.0% (1.0% per year)

Table 16: Estimates of overall screening population and target population for supplemental screening, by age.

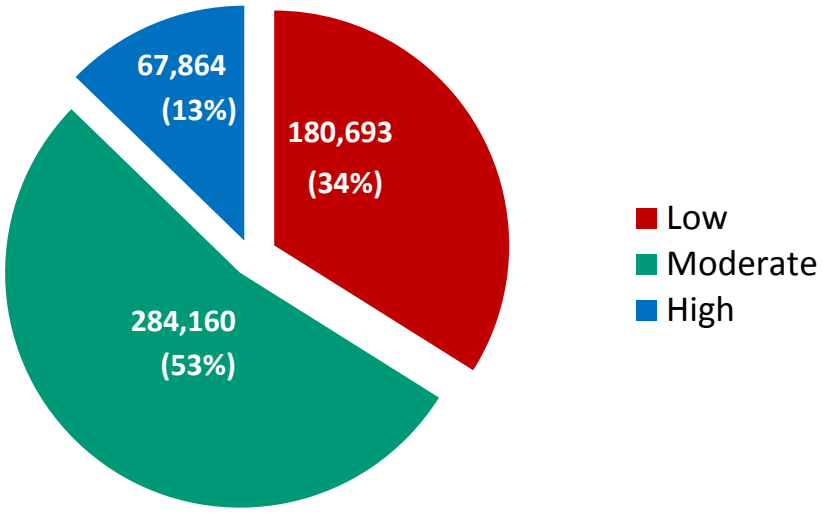
Population/Age	Estimate	Sources
Screening Population, N		2010 U.S. Census (Washington)
40-49	451,211	Whittemore 2004
50-59	458,398	SEER Cancer Statistics 2010
60-69	319,335	Malone, 2006
70-74	93,371	
TOTAL	1,322,315	
Heterogeneously Dense, %		Kerlikowske 2013
40-49	45.6	
50-59	38.6	
60-69	29.0	
70-74	26.0	
Extremely Dense, %		Kerlikowske 2013
40-49	13.7	
50-59	7.5	
60-69	3.8	
70-74	3.1	
Dense Breasts & Close Family History of Breast Cancer, %	22.7	Titus-Ernsthoff, 2006[189]

Support for these thresholds is available in the literature. Studies of chemoprevention generally consider a 5-year risk of approximately 1.7% to be the lower threshold for considering prophylaxis with tamoxifen or other measures to reduce breast cancer risk (Fisher, 1998; Gail, 1999) while the U.S. Preventive Services Task Force's consideration of the same topic categorized women with 5-year risks >3% to be "higher than average risk" (USPSTF, 2002). Our 5-year risk estimate of 5.0% is comparable to the more commonly-used lifetime risk threshold of >20%, which is often listed as a criterion for MRI supplemental screening in a high-risk cohort.

Based on the risk categories described above, we estimate that, of all Washington women with dense breast tissue and a negative digital mammogram, 34% would be low-risk, 53% moderate-risk and 13%

high-risk. These proportions are displayed in Figure 2 below along with the relevant estimated population sizes for each risk group.

Figure 2: Estimated numbers of Washington women with dense breast tissue and negative mammography results, by level of overall breast cancer risk.



Test Diagnostic Performance

We obtained information on each test's diagnostic performance based on data identified during the systematic review. We used published data directly comparing the performance of film vs. digital mammography and the actual prevalence of cancer in the BCSC cohort (Kerlikowske, 2011) and applied these measures to the Washington screening population; wherever possible, density-specific information was used (see Table 17 on the following page). As noted above, there are currently no available DBT studies with long enough follow-up to estimate true population sensitivity. For primary analyses, we assumed a modest (+1.5%) improvement in sensitivity over digital mammography. We derived a specificity estimate of 91.5% based on data on screening volume, recall rates, biopsy rates, and PPV3 from the Friedewald multi-center study (Friedewald, 2014), which happens to also be a 1.5% improvement over the overall specificity estimated in the BCSC digital mammography cohort (Kerlikowske, 2011). Alternative analyses examined (a) a 3 percentage-point improvement in sensitivity and specificity for DBT over digital mammography; (b) a 5 percentage-point improvement; and (c) no improvement in sensitivity and a 1.5 percentage-point improvement in specificity.

Finally, while available DBT studies all suggest that this technology will find additional cancers that would not appear on mammography until the next screening, lack of sufficient follow-up makes such a conclusion premature. We therefore did not assume any incremental cancer detection in primary analyses other than as a product of improved sensitivity; however, we also conducted alternative analyses in which DBT would detect one additional cancer per 1,000 women screened, and examined the impact of this in both our frontline and supplemental screening scenarios. Rates of recall and biopsy were estimated directly from the Friedewald study (Friedewald, 2014), the largest U.S.-based study of DBT published to date. Recall rates in this study were 91.0 and 107.0 per 1,000 screened for DBT and digital mammography respectively; corresponding rates of biopsy were 19.3 and 18.1 per 1,000.

Table 17: Estimates of digital mammography and DBT performance.

Measure	Estimate		Sources
	Digital Mammography	*DBT*	
Sensitivity by BI-RADS Density, %			
Fatty	78.3	79.8	Kerlikowske 2011
Scattered Fibroglandular	86.6	88.1	Assumption
Heterogeneously Dense	82.1	83.6	
Extremely Dense	83.6	85.1	
OVERALL	83.9	85.4	
Specificity by BI-RADS Density, %			
Fatty	94.7	96.2	Kerlikowske 2011
Scattered Fibroglandular	91.2	92.7	Friedewald 2014
Heterogeneously Dense	87.3	88.8	
Extremely Dense	88.7	90.2	
OVERALL	89.8	91.3	
Recall Rate, per 1,000 screened	107.0	91.0	Friedewald 2014
Biopsy Rate, per 1,000 screened	18.1	19.3	Friedewald 2014

BI-RADS: Breast Imaging-Reporting and Data System; DBT: Digital breast tomosynthesis

Unfortunately, there are no available data documenting the performance of MRI in populations with the risk profile assumed for this model. In addition, there are multiple concerns with use of HHUS data as described in this review, including use of film mammography in most studies, lack of complete interval follow-up, and heterogeneity of study populations. We opted to estimate rates of detection of cancers missed on mammography (i.e., interval cancers) based on an extrapolation of the relative sensitivity of HHUS to MRI in similar populations (see Table 18 on the following page). Specifically, we assumed that HHUS and MRI would detect 80% and 95% of cancers respectively when all risk groups are considered. In the absence of data, we assumed that all measures of ABUS performance would be equivalent to that of HHUS, and the two modalities would differ only in terms of cost.

Estimates of supplemental screening performance can be found in Table 18 on the following page. The rate of interval cancers in most screening populations is approximately 1 per 1,000 (Kerlikowske, 2011; NHSBSP, 2012). However, ultrasound and MRI also detect cancers that would not have presented clinically during the interval between frontline screenings, particularly in populations at higher-than-average breast cancer risk. Evidence from this review suggests that the rate of incremental cancer detection in relevant HHUS studies is approximately 3 per 1,000. Because data from available MRI studies were representative of very high-risk women only, cancer detection rates from these studies would represent an overestimate in our target population. We therefore conservatively assumed that MRI would identify an additional 5 cases of cancer per 1,000.

These studies were also assessed for information on cancer yield per biopsy. We assumed a rate of 8.5% for HHUS/ABUS based on a mean of values from the most representative studies identified. MRI's positive predictive value was 25% among women recommended for biopsy in a recent meta-analysis (Warner, 2008); however, because the studies included in this review were conducted exclusively in high-risk women, we reduced our estimate to 20% to reflect MRI's potential use in a mixed-risk population. Cancer detection rates were divided by cancer yield estimates to obtain an estimate of the total number of biopsies performed for each modality.

Table 18: Estimates of supplemental screening performance.

Measure/Test	Risk Level				Sources
	Low	Moderate	High	Overall	
Sensitivity for Interval Cancers, %					
HHUS/ABUS	75.0	80.0	85.0	80.0	Review
MRI	87.5	95.0	100.0	95.0	Review
Additional Cancers Detected (per 1,000)					
HHUS/ABUS	1.5	3.5	5.0	3.0	Review
MRI	3.0	5.5	8.5	5.0	Review
Cancer Yield per Biopsy, %					
HHUS/ABUS	7.3	8.5	9.5	8.5	Review
MRI	15.0	20.0	25.0	20.0	Review; Warner, 2008

NOTE: ABUS performance assumed to be equivalent to HHUS

HHUS: Handheld ultrasound; ABUS: Automated breast ultrasound; MRI: magnetic resonance imaging; DBT: digital breast

While sensitivity and specificity are considered to be independent of the prevalence of underlying disease, prevalence is often a marker for clinical variability, and measures of diagnostic accuracy often improve with increasing prevalence (Leeflang, 2009). We therefore assumed that all of these measures (i.e., detection of interval and additional cancers, and cancer yield per biopsy) would increase with increasing levels of breast cancer risk. We developed estimates for each risk subgroup that would equate to the overall levels described above when weighted by population size.

Definitions and estimates of the proportion of "overdiagnosed" cancers (i.e., those detected that would never have otherwise required treatment) vary substantially across studies. We estimated that between 10-30% of biopsy-detected cancers would be cases of overdiagnosis, based on the range generally reported in the literature (Welch, 2010; de Gelder, 2011; Zackrisson, 2006; Independent UKPoBCS, 2012). We present the lower and upper boundaries of this range for each screening scenario evaluated.

Testing and Biopsy Costs

We adopted a third-party payer perspective for the model, and therefore focused on estimates of payment for screening, diagnostic imaging, and biopsy (see Table 19 below). Payments were estimated primarily using the 2014 Medicare fee schedule (Centers for Medicare and Medicaid Services, 2013).

There is currently no separate and standardized reimbursement coding for ABUS. However, use of "add-on" codes has been reported for ABUS (GE Healthcare, 2013), which are reflected in our estimates. At the time of publication of the draft report, CMS had not yet ruled on new codes and payment for DBT. Payment for DBT was based on the rate for digital mammography with an additional fee charged to the patient; we used an estimate of $50 in primary analyses based on the lower end of the reported range from facility websites (University of Virginia Health System, 2013; South Jersey Radiology, 2013; Washington Radiology Associates, 2013). CMS published a final rule on October 31, 2014 that established billing codes and payment rates for DBT beginning in January 2015; the estimated national payment rate for DBT will be approximately $57 above that for digital mammography (ACR, 2014). We used this amount but also varied DBT costs according to the potential premium that might be paid by a variety of insurers ($10, $25, $50, and $75); we also allowed for the possibility that the previous standard would hold (i.e., no premium in reimbursement over mammography).

The costs of diagnostic workup for women recalled after positive mammography included those of a unilateral diagnostic mammogram in all and an HHUS exam in 50%, as well as biopsy in those so referred. We assumed that women presenting with an interval (i.e., "missed") cancer would present clinically and receive both a unilateral diagnostic mammogram and a biopsy. For each supplemental screening strategy, costs of interest included those of screening, biopsy, and diagnosis of interval cancers.

Table 19: Payment estimates for mammography, biopsy, and supplemental screening modalities.

Test/Procedure	Components	Payment	Source(s)
Digital Mammography	Bilateral exam, computer-aided detection	$145.44	Medicare fee schedule

Test/Procedure	Components	Payment	Source(s)
HHUS	Bilateral breast ultrasound	$99.95	Medicare fee schedule
ABUS	HHUS, +3D rendering	$184.13	Medicare fee schedule
MRI	Bilateral breast MRI, computer-aided detection	$562.06	Medicare fee schedule
DBT	Bilateral exam, computer-aided detection+additional views	$202.01	Medicare fee schedule, plus $56.57 expected additional payment for 2015
Diagnostic Mammography (Digital)	Unilateral exam, computer-aided detection	$140.07	Medicare fee schedule
Biopsy	Biopsy with ultrasound or stereotactic guidance, surgical biopsy	$1,303.40	Medicare fee schedule (assumes 75% percutaneous, 25% surgical)

HHUS: Handheld ultrasound; ABUS: Automated breast ultrasound; MRI: magnetic resonance imaging; DBT: digital breast tomosynthesis

The costs of diagnostic workup for women recalled after positive mammography included those of a unilateral diagnostic mammogram in all and an HHUS exam in 50%, as well as biopsy in those so referred. We assumed that women presenting with an interval (i.e., "missed") cancer would present clinically and receive both a unilateral diagnostic mammogram and a biopsy. For each supplemental screening strategy, costs of interest included those of screening, biopsy, and diagnosis of interval cancers.

9.3 Model Results

Population Estimates

As mentioned previously, 47% of the 1.3 million women age 40-74 in Washington expected to undergo mammography screening would have BI-RADS density "c" or "d" (620,000). Of these women with dense breasts, 86% (533,000) would be expected to have a negative digital mammogram and therefore be candidates for supplemental screening. Use of DBT as the frontline screening modality would increase the candidate population slightly (to 542,000) as a result of improved specificity – in other words, fewer

DBT-screened women would be recalled for further testing and biopsy, and instead be classified as having "negative" or "normal" results.

Comparison of DBT vs. Digital Mammography

The expected performance of DBT vs. digital mammography in Washington is compared in Table 20 on the following page for the overall screened population as well as the subset of women with dense breast tissue. To facilitate comparisons, we present all clinical findings on a "per 1,000 women screened" basis, and costs are presented as an average per woman screened.

As shown in the table, DBT results in a small increase in the number of cancers detected (3.7 vs. 3.6 per 1,000 for digital) and a small decrease in the number of cancers missed (0.6 vs. 0.7 per 1,000) when compared to digital mammography for the overall screening population. Rates of false-positive results with or without biopsy were both lower for DBT, owing to its slightly better specificity and reduced recall rate relative to digital mammography. However, screening costs were *not* offset by reduced levels of diagnostic workup. Total costs per woman screened were $56 higher for DBT ($245 vs. $189 for digital mammography), meaning that only 2% of the $57 premium in screening cost for DBT would be offset by reduced levels of unnecessary diagnostic workup. Because the estimated costs of recall without biopsy are <15% of the costs of recall with biopsy (i.e., $190 vs. $1,493 respectively), most of the savings from reduced recall are washed out by increased costs of biopsy with DBT.

Not surprisingly, recall and biopsy rates were higher in the subset of women with BI-RADS "c" or "d" breast density, as the incidence of cancer was higher with increasing breast density in the BCSC cohort. For example, cancer occurred at a rate of approximately 5 per 1,000 in women with extremely dense breasts, vs. 2 per 1,000 in those with fatty breasts (BI-RADS 1).

Table 20: Clinical outcomes and costs of general population breast cancer screening in Washington: comparison of digital mammography vs. digital breast tomosynthesis.

Outcome (per 1,000 screened)	Digital Mammography	DBT
Overall Population		
Recalls	107.0	91.0
Biopsies Performed	18.1	19.3
Cancers Detected (True Positives)	3.6	3.7
False Positive (with Biopsy)	14.5	15.6
False Positive (without Biopsy)	83.3	67.2
Cancers Missed (Interval Cancers)	0.7	0.6
Cost (per Woman Screened, $)	189	245
Women w/Dense Breast Tissue		
Recalls	130.6	114.6
Biopsies Performed	22.1	24.3
Cancers Detected (True Positives)	4.2	4.3
False Positive (with Biopsy)	17.9	20.0
False Positive (without Biopsy)	105.7	89.6

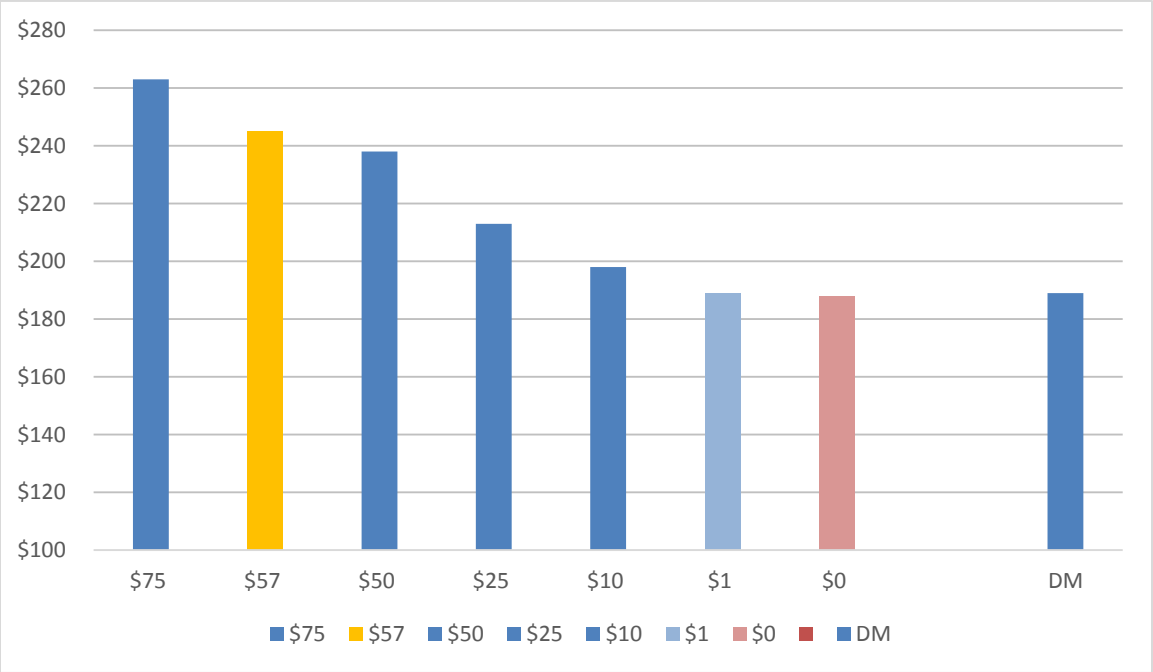
Outcome (per 1,000 screened)	Digital Mammography	DBT
Cancers Missed (Interval Cancers)	0.9	0.8
Cost (per Woman Screened, $)	194	249

NOTES: Recalls refer to positive mammograms or DBTs recalled for additional imaging and/or biopsy; findings may not sum perfectly due to rounding

Among women with dense breast tissue, absolute levels of recalls and biopsies performed were higher in comparison to the overall cohort. However, because no density-specific differences in diagnostic performance were assumed for DBT relative to digital mammography, incremental differences were the same as in the overall cohort (i.e., differences of 0.1 per 1,000 in both cancers detected and cancers missed).

Findings for our sensitivity analyses are presented in Figure 3 below Table 21 on the following page. The basecase cost for DBT is shown on the yellow bar. Variations in the dollar premium added for DBT affected only screening costs (i.e., performance statistics remained the same). At a premium of $75, the total cost per woman screened with DBT was $263, or a 39% increase over the cost of digital mammography alone. In contrast, at a small premium of $10, total costs per woman screened were increased by only 5%. A DBT-based screening strategy would be cost neutral to that of digital mammography at a $1 premium (as shown on the light blue bar), and would begin to be cost-saving if the previous Medicare reimbursement approach (i.e., no additional payment for DBT) were to hold (as shown on the light red bar).

Figure 3. Total cost per woman screened, at different payment premiums for digital breast tomosynthesis over digital mammography.



Results of our analyses varying the test performance of DBT can be found in Table 21 on the following page. Importantly, the budgetary impact of DBT at a $57 premium remained substantial, even in the

most optimistic of scenarios regarding diagnostic accuracy. When DBT was assumed to improve only specificity over digital mammography (scenario A), the only parameters to change from primary analyses related to small changes in the numbers of cancers detected and missed. Costs did not materially change from the basecase. More substantial improvements in both sensitivity and specificity were modeled in scenarios B and C. For example, when DBT was assumed to have 87% sensitivity and 93% specificity (scenario B), the recall rate declined by 21% (71.6 vs. 91.0 per 1,000 in primary analyses), and the biopsy rate also declined substantially (15.2 vs. 19.3 per 1,000). The cost per woman screened was estimated to be $242 under these assumptions, meaning that 7% of the additional $57 in DBT payments would be offset by fewer false-positive recalls and biopsies. Scenario C assumed an 89% sensitivity and 95% specificity for DBT; in this analysis, the recall rate would be reduced by over 40% relative to the basecase (51.7 vs. 91.0 per 1,000); the biopsy rate would be 11 per 1,000 (vs. 19.3 per 1,000 in the basecase). The cost per woman screened would further decline to $238, meaning that 14% of the additional payments for DBT would be offset by improved test performance in this scenario. It should be noted, however, that a specificity level this high was only reported in the Norwegian study (Skaane, 2013); as discussed, the process of adjudicating breast images is more intense than in the U.S., and 95% specificity may not be achievable here. The final scenario assumed that basecase estimates for sensitivity and specificity (85.5% and 91.5% respectively) would apply, but that DBT would detect an additional 1 cancer per 1,000 screened that would not have been detectable between mammography screenings. The only changes were to the cancer detection rate (4.7 vs. 3.7 per 1,000 in the basecase) and the false-positive biopsy rate (14.6 vs. 15.6 per 1,000). Costs were essentially unchanged in this scenario.

Table 21. Results of sensitivity analyses varying test performance of digital breast tomosynthesis and number of additional cancers detected vs. digital mammography.

Outcome (per 1,000 screened)	DM Basecase Sn: 84.0 Sp: 90.0	DBT Basecase Sn: 85.5 Sp: 91.5	(A) Sn: 84.0 Sp: 91.5	(B) Sn: 87.0 Sp: 93.0	(C) Sn: 89.0 Sp: 95.0	(D) 1 add'l cancer detected
Overall Population						
Recalls	107.0	91.0	91.0	71.6	51.7	91.0
Biopsies Performed	18.1	19.3	19.3	15.2	11.0	19.3
Cancers Detected	3.6	3.7	3.6	3.8	3.9	4.7
False + (with Biopsy)	14.5	15.6	15.6	11.4	7.1	14.6
False + (w/o Biopsy)	83.3	67.2	67.2	52.3	32.4	67.2
Interval Cancers	0.7	0.6	0.7	0.6	0.5	0.6
Cost (per Woman Screened)	$189	$245	$245	$242	$238	$244

Sn: Sensitivity; Sp: Specificity; DM: Digital mammography; DBT: Digital breast tomosynthesis

NOTES: Recalls refer to positive mammograms or DBTs recalled for additional imaging and/or biopsy; findings may not sum perfectly due to rounding

Incremental Effects of Supplemental Screening in Women with Dense Breast Tissue and a Negative Digital Mammogram

We compared the three supplemental screening scenarios (HHUS, ABUS, MRI, and DBT) to no supplemental screening (i.e., digital mammography alone) on an overall basis as well as separately for low, moderate, and high-risk women. Results are described for each group of interest in the sections that follow.

Overall (All Risk Groups Combined)

Findings for the combined population of low-, moderate-, and high-risk women can be found in Table 22 on the following page. As discussed previously, neither recalls nor false-positives without biopsy were estimated for these analyses, as all positive supplemental screening results were assumed to result in biopsy. We present clinical results for HHUS or ABUS together, as equivalent performance was assumed. Costs were assumed to differ, however, and are presented separately at the bottom of the table.

The addition of MRI to digital mammography detects more cancers (6.0 vs. 3.8 for HHUS/ABUS). HHUS/ABUS would nearly quadruple the number of biopsies required over digital mammography alone, while biopsies would increase nearly threefold with MRI. Each of the supplemental modalities would identify nearly all of the cancers missed by mammography. MRI was the more costly strategy ($602), however, due to the higher payment rate for the test itself. Costs were $159 and $243 for HHUS and ABUS respectively.

Table 22: One-year clinical outcomes and costs of supplemental screening in Washington in all women with dense breast tissue and negative mammography: vs. digital mammography alone.

Outcome (per 1,000 screened)	DM+HHUS/ABUS	DM+MRI	DM Alone
Biopsies Performed	67.9	52.6	22.1
Incremental increase	*45.0*	*30.4*	
Cancers Detected (True Positives	8.0	10.2	4.2
Incremental increase	*3.8*	*6.0*	
Adjusted for potential overdiagnosis (low)	*3.4*	*5.4*	
Adjusted for potential overdiagnosis (high)	*2.6*	*4.2*	
False Positive Biopsy	59.1	42.4	17.9
Incremental increase	*41.2*	*24.5*	
Cancers Missed (Interval Cancers)	0.2	0.1	0.9
Incremental improvement	*(0.7)*	*(0.8)*	
Cost (per Woman Screened, $)	353/437	796	194
Incremental increase	*159/243*	*602*	

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

Low Risk

Clinical and economic outcomes of supplemental screening in low-risk (i.e., 5-year risk of 1%) women are presented in Table 23 on the following page. In this low-risk population, HHUS and ABUS identify a relatively small number of additional cancers (1.8 per 1,000), while MRI detects an incremental 3.4 per

1,000 screened. The number of biopsies required to detect this small number of cancers with MRI or HHUS is approximately five times higher than the rate for DM alone, however. In this low-risk population, seven and 14 biopsies would be required for MRI and HHUS/ABUS respectively to detect one additional cancer over digital mammography alone. As in the overall population, each of the supplemental strategies would detect nearly all of the cancers missed by mammography.

All supplemental strategies would substantially increase screening costs compared with DM alone. Use of HHUS would increase costs by $133 (~70%) per woman screened, while the assumed greater expense for ABUS would nearly double screening costs. Screening costs would increase over fourfold compared to DM alone with the addition of MRI to DM.

Table 23: Clinical outcomes and costs of supplemental screening in Washington in women at low overall breast cancer risk with dense breast tissue and negative mammography: vs. digital mammography alone.

Outcome (per 1,000 screened)	DM+HHUS/ABUS	DM+MRI	DM Alone
Biopsies Performed	32.8	30.2	7.7
Incremental increase	*25.1*	*22.5*	
Cancers Detected (True Positives	3.4	5.0	1.6
Incremental increase	*1.8*	*3.4*	
Adjusted for potential overdiagnosis (low)	*1.6*	*3.0*	
Adjusted for potential overdiagnosis (high)	*1.2*	*2.4*	
False Positive Biopsy	29.4	25.2	6.1
Incremental increase	*23.3*	*19.1*	
Cancers Missed (Interval Cancers)	0.1	0.1	0.4
Incremental improvement	*(0.3)*	*(0.3)*	
Cost (per Woman Screened, $)	321/405	779	188
Incremental increase	*133/217*	*591*	

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

Moderate Risk

Findings for patients in the moderate-risk group (5-year risk of 2.5%) can be found in Table 24 on the following page. The higher prevalence of cancer in this subgroup is associated with higher rates of biopsy and false-positive results for all tests. HHUS/ABUS would detect an additional 4.4 cancers per 1,000 women screened, while MRI would detect 6.5. The numbers of biopsies required to detect an additional cancer were five and 12 for MRI and HHUS respectively. There was slight separation in the number of interval cancers that would have been missed by supplemental screening (0.1 for MRI vs. 0.2 for HHUS/ABUS). Differences in cost were similar to those observed in the low-risk subgroup.

Table 24: Clinical outcomes and costs of supplemental screening in Washington in women at moderate overall breast cancer risk with dense breast tissue and negative mammography: vs. digital mammography alone.

Outcome (per 1,000 screened)	DM+HHUS/ABUS	DM+MRI	DM Alone
Biopsies Performed	70.5	51.8	19.2
Incremental increase	*51.3*	*32.6*	
Cancers Detected (True Positives	8.3	10.4	3.9
Incremental increase	*4.4*	*6.5*	
Adjusted for potential overdiagnosis (low)	*4.0*	*5.8*	
Adjusted for potential overdiagnosis (high)	*3.1*	*4.5*	
False Positive Biopsy	62.2	41.4	15.3
Incremental increase	*46.9*	*26.1*	
Cancers Missed (Interval Cancers)	0.2	0.1	1.1
Incremental improvement	*(0.9)*	*(1.0)*	
Cost (per Woman Screened, $)	363/447	801	196
Incremental increase	*167/251*	*605*	

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

High Risk

Outcomes and costs of supplemental screening for women at high risk of breast cancer (5-year risk of 5%) are presented in Table 25 on the following page. Greater than 10% of women would undergo a biopsy after screening with mammography and HHUS; over two-thirds of biopsies would come from the HHUS component of screening. Estimated totals of four and 11 biopsies would be required for MRI and HHUS/ABUS to detect each additional case of cancer. MRI would correctly identify all cancers in high-risk women, although approximately one to three of the approximately 11 cancers identified have the potential to be cases of overdiagnosis. In addition, MRI would miss none of the cancers that would have been missed on mammography, while HHUS/ABUS would miss 0.3 cases per 1,000. Differences in false-positive rates are also magnified in the high-risk population. HHUS would produce a rate of false-positive biopsies more than twice that of MRI (65.0 vs. 31.9 per 1,000 respectively). MRI remained the more costly supplemental test strategy of the three modalities; including costs of mammography, an MRI-based strategy would cost over $800 per woman screened in the high-risk group.

Table 25: Clinical outcomes and costs of supplemental screening in Washington in women at high overall breast cancer risk with dense breast tissue and negative mammography: vs. digital mammography alone.

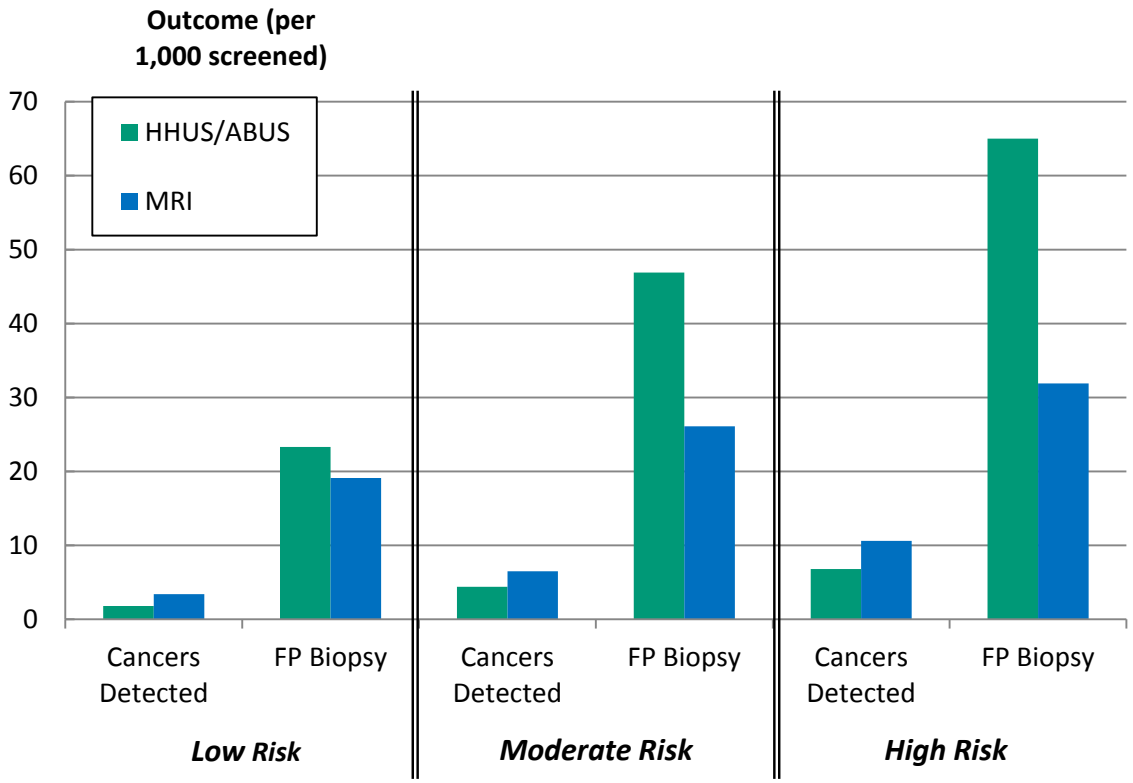
Outcome (per 1,000 screened)	DM+HHUS/ABUS	DM+MRI	DM Alone
Biopsies Performed	110.4	81.2	38.6
Incremental increase	*71.8*	*42.6*	
Cancers Detected (True Positives	14.7	18.5	7.9
Incremental increase	*6.8*	*10.6*	
Adjusted for potential overdiagnosis (low)	*6.1*	*9.5*	
Adjusted for potential overdiagnosis (high)	*4.7*	*7.4*	
False Positive Biopsy	95.7	62.6	30.7
Incremental increase	*65.0*	*31.9*	
Cancers Missed (Interval Cancers)	0.3	---	2.1
Incremental improvement	*(1.8)*	*(2.1)*	
Cost (per Woman Screened, $)	396/480	820	202
Incremental increase	*194/278*	*618*	

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

Comparison of Risk Groups

Key incremental effects of supplemental screening (i.e., above and beyond effects of digital mammography alone) are compared for each risk group in Figure 4 on the following page. While differences between modalities in the number of additional cancers detected remain relatively stable with increasing risk, differences in rates of false-positive biopsy become more pronounced. For example, HHUS/ABUS would produce 4.2 more false-positive biopsies per 1,000 women screened than MRI in low-risk women (23.3 vs. 19.1 per 1,000 respectively), but would generate 33.1 more per 1,000 among those in the high-risk group (65.0 vs. 31.9 per 1,000).

Figure 4: Selected incremental effects of supplemental screening, by screening modality and overall breast cancer risk.



HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; FP: False positive

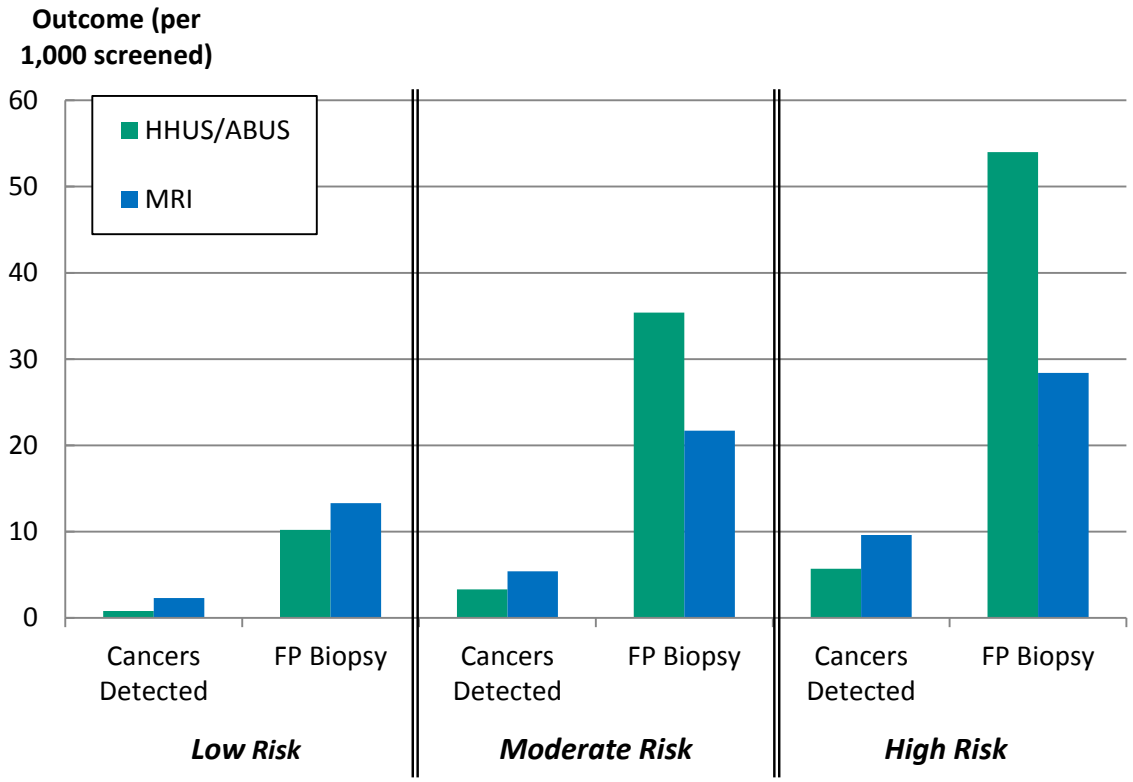
Incremental Effects of Supplemental Screening in Women with Dense Breast Tissue and a Negative Digital Breast Tomosynthesis

We also estimated the effects of supplemental screening with HHUS/ABUS and MRI if frontline screening were performed with DBT rather than digital mammography. Under our basecase analysis, the small improvement in sensitivity with DBT would mean that slightly fewer cancers (0.1 per 1,000) would be missed and available for detection by supplemental screening. The small improvement in specificity would translate into fewer false positives on routine screening, which would have the effect of sending approximately 2% more women with dense breast tissue into supplemental screening (see Section 9.4 for further information on population impact). Even with this small increase, however, rates of biopsy and false-positive biopsy would not appreciably change on a per-1,000 screened basis, nor would costs of supplemental screening.

When we assumed that frontline screening with DBT would also detect one additional cancer per 1,000 that would not be detectable by mammography, however, a different picture emerged. Figure 5 on the following page presents the numbers of cancers detected and false-positive biopsy rates for women at low, moderate, and high overall breast cancer risk. Rates of cancer detection did not change appreciably, as there was only one fewer cancer to detect per 1,000 women screened. However, biopsy

rates (and false-positive rates) declined sharply as a result of DBT's improved performance. For example, among higher-risk women, the false-positive biopsy rate for HHUS/ABUS declined from 65 per 1,000 screened under basecase assumptions to 54 per 1,000 in this scenario. Interestingly, the false-positive biopsy rate was lower for HHUS/ABUS than for MRI among low-risk women in this analysis, as nearly three times as many cancers would be detectable by MRI vs. HHUS/ABUS in this subgroup (2.4 vs. 0.9 per 1,000).

Figure 5: Selected incremental effects of supplemental screening, by screening modality and overall breast cancer risk: when DBT is frontline test and is assumed to detect 1 additional cancer per 1,000.



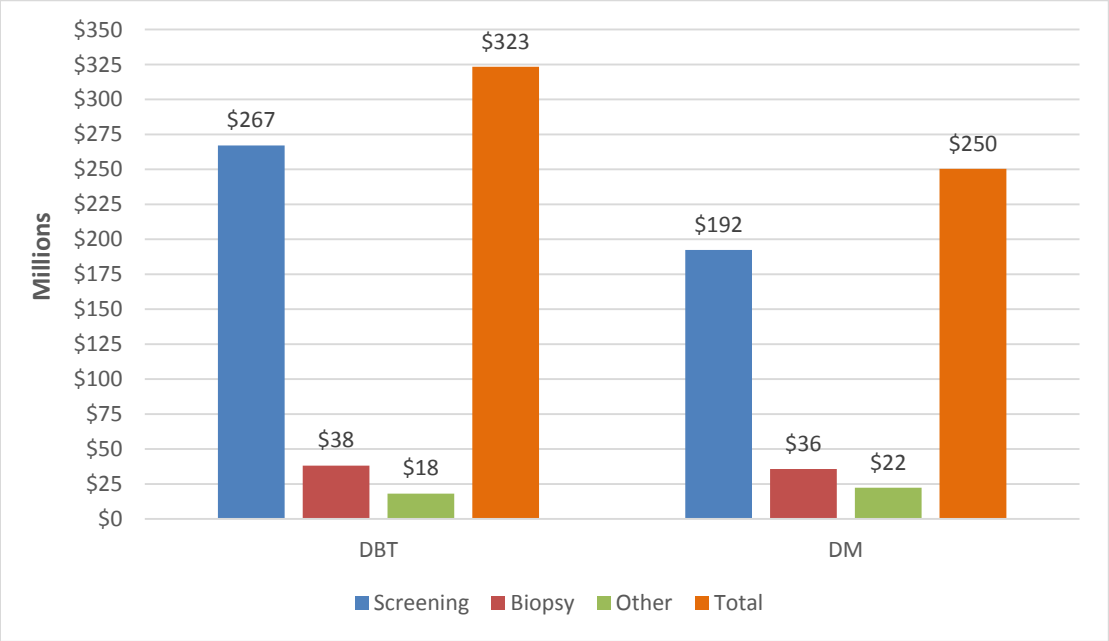
HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; FP: False positive

Also, while not presented in Figure 5 above, costs also declined as a result of fewer biopsies. Across all risk groups, the costs per woman screened with HHUS, ABUS, and MRI in this analysis were $142, $226, and $594 respectively; corresponding values in the basecase analysis were $159, $243, and $602. On a relative basis, the decline in costs was greatest for HHUS (11%, vs. 7% for ABUS and 2% for MRI), as the costs of the screening test itself made up a greater proportion of total costs for ABUS and MRI.

9.4 Population Impact of Frontline and Supplemental Screening

We also estimated the impact of our modeled estimates of frontline and supplemental screening performance and costs when applied to the screening-eligible population in the state of Washington. On a population basis, the costs of frontline screening with digital mammography, including costs of screening, diagnostic workup, and biopsy, are estimated to total approximately $250 million annually. Replacement of digital mammography with DBT in all women would raise costs by nearly 30%, to $323 million annually, using our basecase estimates of +1.5% absolute improvement in both sensitivity and specificity and a $57 premium on screening payment. As shown in Figure 6 below, savings from reduced recalls with DBT are diminished by higher biopsy costs, so the difference in cost comes almost entirely from the increased cost of the screening test itself.

Figure 6: Population-based estimates of the cost of routine breast cancer screening among eligible women age 40-74 in the state of Washington.



DBT: Digital breast tomosynthesis; DM: Digital mammography

NOTE: "Other" includes cost of diagnostic workup for recalls not involving biopsy and costs of diagnosis for women presenting with interval cancers

As described previously, use of supplemental screening technologies would incrementally add to the numbers of cancers detected but also biopsies performed among women with negative initial screening result. Earlier findings were presented on a per 1,000 screened basis. Population-based estimates of the incremental clinical impact of supplemental screening with HHUS/ABUS and MRI among all women with dense breast tissue in the state of Washington can be found in Table 26 on the following page.

Table 26: Population-based estimates of incremental clinical impact of supplemental screening among Washington women with dense breast tissue and negative mammography results, by supplemental screening modality and overall breast cancer risk.

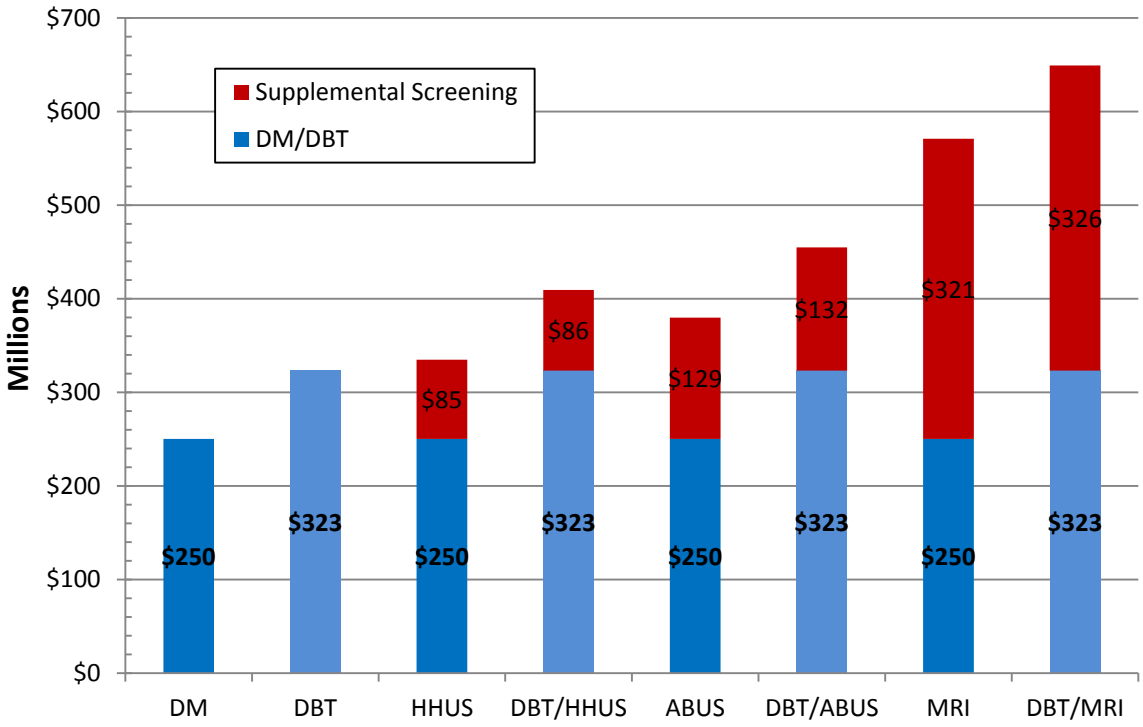
Outcome/Cost	HHUS/ABUS	MRI
Low Risk (n=180,693)		
Biopsies Performed	4,539	4,065
Cancers Detected (True Positives)	329	610
False Positive Biopsies	4,209	3,455
Cancers Missed (Interval Cancers)	19	10
Moderate Risk (n=284,160)		
Biopsies Performed	14,562	9,259
Cancers Detected (True Positives)	1,238	1,852
False Positive Biopsies	13,325	7,407
Cancers Missed (Interval Cancers)	61	15
High Risk (n=67,864)		
Biopsies Performed	4,871	2,888
Cancers Detected (True Positives)	463	722
False Positive Biopsies	4,408	2,166
Cancers Missed (Interval Cancers)	22	---

DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging

These results highlight the differences in clinical tradeoffs when supplemental screening strategies are used in the low- and high-risk populations. For example, the numbers of total and false-positive biopsies would be similar in these two subgroups, but supplemental screening would detect over 100 more cancers in the high-risk population despite the fact that it is 40% the size of the low-risk group.

The estimated budgetary impact to Washington of supplemental screening in all women with dense breasts and negative mammography can be found in Figure 7 on the following page. As mentioned above, the annual cost of digital mammography screening is estimated to total approximately $250 million in the state of Washington. Supplemental screening of all women with an initial negative digital mammogram with HHUS would increase annual costs by approximately 35%, to $335 million. Use of higher-cost ABUS as the modality of choice would result in an increase of over 50% in costs (to $380 million) for the same assumed clinical benefit. Finally, use of MRI results in a more than twofold increase in overall costs (to over $570 million) annually.

Figure 7: Costs of digital mammography and supplemental screening among all Washington women with dense breast tissue and negative frontline screening result, by screening modality.



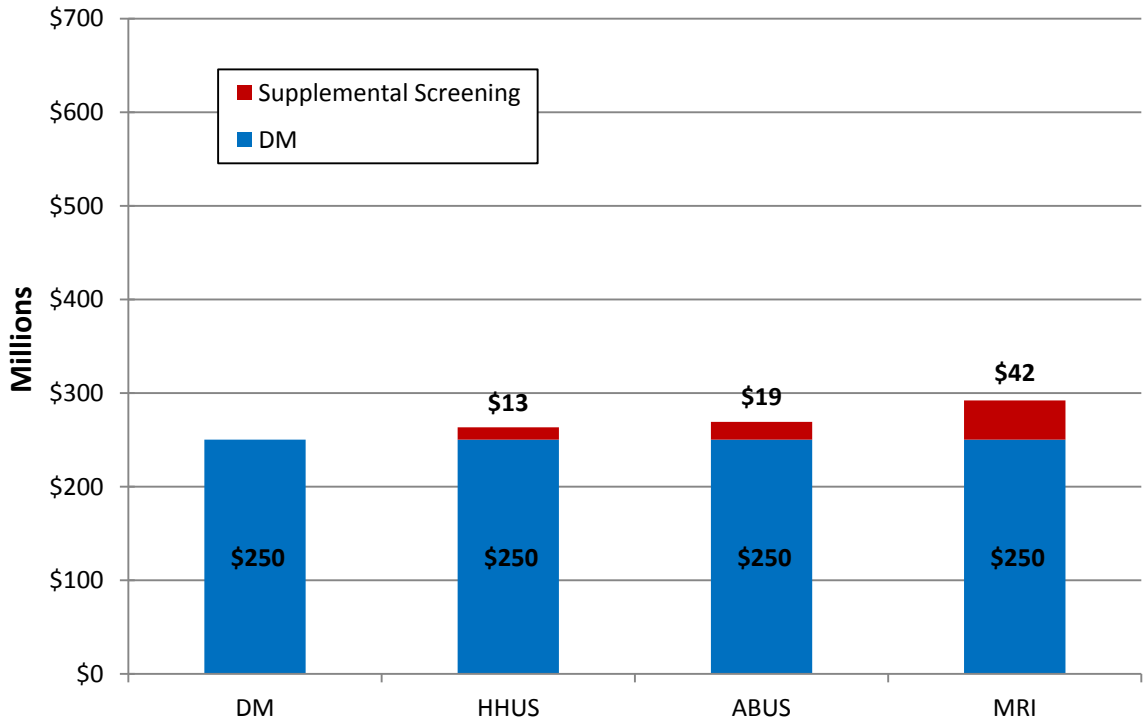
DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; DBT: Digital breast tomosynthesis

When DBT is used as the general population screening modality, total costs of screening are estimated to be $323 million. While most of this increase is due to increased screening costs, costs of supplemental screening also increase slightly (by $1-$6 million depending on the modality used) in this scenario, given the increased size of the population available for supplemental screening. Specifically, approximately 9,000 more women would be screened after a negative DBT; these women would have been recalled for further imaging after digital mammography.

Also, while not represented in the figure above, we examined the impact on total screening costs under the scenario in which DBT detected one additional cancer per 1,000 woman screened that would not have been detectable on mammography. While frontline screening costs did not materially change in this scenario, supplemental screening costs were lower as a result of fewer biopsies required to detect available cancers. Specifically, supplemental screening costs declined by $9 million for HHUS and ABUS, and by $4 million for MRI relative to the primary DBT analysis (i.e., improved sensitivity and specificity, but no additional cancer detection benefit).

Regardless of the frontline screening modality employed, a substantial proportion of the additional costs of supplemental screening are generated in the low-risk population, the subgroup in which the fewest additional cancers are detected. Figure 8 below shows the additional costs of supplemental screening after negative digital mammography when limited to women in the "higher risk" category.

Figure 8: Costs of digital mammography and supplemental screening among "*high-risk*" Washington women with dense breast tissue and negative mammography, by screening modality.



DM: Digital mammography; HHUS: Handheld ultrasound; ABUS: automated breast ultrasound; MRI: magnetic resonance imaging; DBT: digital breast tomosynthesis

If supplemental screening were limited to women age 50-74 with dense breast tissue, a family history in a first degree relative, and a negative digital mammogram (i.e., the high-risk cohort), total costs of screening would rise by a much smaller increment. However, the potential yield of additional cancers detected in this subgroup would be comparable to or better than with digital mammography alone. For example, supplemental MRI screening in high-risk women would increase costs by approximately $42 million (17%) to approximately $290 million, and would find a total of 1,358 cases of cancer (636 cancers from digital mammography alone + 722 additional cancers from MRI).

Increases in cost would be lower with HHUS and ABUS (5% and 8% respectively), but the additional cancer yield would also be lower (463 additional cancers detected over digital mammography alone). Findings such as these are important to consider in any evaluation of the tradeoffs of supplemental screening, including numbers of biopsies required, additional cancers detected and missed, and screening costs.

9.5 Model Limitations

We note important limitations of this cohort model. First, as required by any modeling approach, we made a number of simplifying assumptions that may not truly reflect the use of mammography, DBT, or supplemental screening in clinical practice. These included screening behaviors and clinical decisions such as perfect compliance with both types of screening as well as referral for and performance of biopsy in 100% of women with positive supplemental screening results. These assumptions likely resulted in overestimates of rates of cancer detection and cost for mammography, DBT, and supplemental screening.

Our most important assumption, however, was that each supplemental modality would identify additional cancers that would not have presented during the interval between mammography screenings, as has been demonstrated in the studies of interest for this review. However, these modalities have by and large not been studied exclusively in women with dense breast tissue and negative mammography who are at varying levels of overall breast cancer risk. It may be the case, for example, that we have overestimated the performance of HHUS/ABUS in high-risk individuals, as nearly all of the ultrasound studies evaluated in this review have been in women with breast cancer prevalence levels well below 1%. Conversely, we may also have overestimated MRI's performance in low-risk women, as the evidence base for supplemental MRI screening is currently limited to women at very high overall breast cancer risk.

We also included DBT in our analysis as its overall evidence base is emerging. As such, we had to make important assumptions regarding its performance relative to digital mammography in the absence of sufficient follow-up data in available studies. In addition, while CMS has recently calculated an incremental payment rate for DBT above that for digital mammography alone ($57), the true cost to *all* third-party payers remains unknown. We varied all of these parameters in sensitivity analyses, but a clear understanding of DBT's benefits will only come from further study with long-term follow-up.

Because the model adopted a payer perspective, we did not measure certain impacts of screening, such as potentially improved screening "throughput" with ABUS over HHUS as well as patient-time costs associated with each modality. Finally, while we attempted to provide reference figures for the number of cancers that might be "overdiagnosed" by these supplemental modalities, this did not explicitly consider the possibility that some proportion of cancers diagnosed by ultrasound or MRI would also have been diagnosed during the next round of mammography or DBT screening. This type of

information will only be available through the conduct of longer-term randomized controlled trials or cohort studies comparing the benefits of supplemental screening to digital mammography or DBT.

9.6 Summary

Despite these limitations, we believe that our model results have important implications for discussion. First, modest improvements in sensitivity and specificity for DBT over digital mammography, resulting in reduced recall and better cancer detection, will likely be accompanied by substantial increases in frontline screening costs, which will exceed reduced downstream costs associated with recall. If these modest improvements also come with an increase in the cancer detection rate, however, this will have the potential to decrease supplemental screening costs while holding frontline DBT costs relatively constant.

With regard to supplemental screening, the use of any screening modality among all women with dense breast tissue and a negative mammogram or DBT has the potential to significantly increase costs to the state due not only to screening costs but to those of false-positive biopsy; such costs would more than double if MRI were used as the supplemental modality, for example. Prioritization of supplemental screening for women at higher overall breast cancer risk would increase costs at a more moderate level (5-17%) while still detecting a substantial number of additional cancers. This also underscores the importance of conversation around breast cancer risk between a woman and her physician – women with dense breast tissue should understand these risks in the context of their age, family history, and other clinical factors, and should be encouraged to seek formal risk assessment and genetic counseling.

10. Recommendations for Future Research

Technological advances in breast cancer screening have come during an era in which the frequency of such screening (and for some, whether screening should be done at all) has come under intense scrutiny. Since the publication of decades-old pivotal trials demonstrating a mortality benefit from mammography, relatively little insight has been gained into the incremental gains in morbidity and mortality that technological innovation can provide, and whether evidence of such gains might be confounded by improvements in breast cancer treatment.

It is this setting in which we must evaluate the evidence on both digital breast tomosynthesis as a frontline screening tool as well as supplemental screening with handheld ultrasound, automated ultrasound, and MRI in women with dense breast tissue and a negative frontline screening result. There are concerns with the evidence on each of these modalities, ranging from lack of adequate follow-up in the initial DBT studies, to evaluation of HHUS and MRI in populations not necessarily representative of all women with dense breast tissue, to the small and emerging evidence base with ABUS and the associated variability in study findings. In all of these situations, however, any correlation between improved test performance and long-term clinical benefit is highly speculative.

There are renewed calls for large, robust, government-funded clinical trials to address this question, including from many advocates of the technologies included in this review. Whether the funds or resources necessary to conduct these trials will be made available is an open question. In the meantime, we feel that several research topics could be prioritized to further the field and address important evidence gaps. For one, simply extending the follow-up duration of new or existing registry-based studies will capture important information on interval cancers, the relative performance of different screening strategies (e.g., biennial vs. annual, age- and/or density-targeted vs. general), and other important outcomes.

In addition, the controversy around overdiagnosis and overtreatment stems in large part from a lack of knowledge of the tumor types that are likely to be aggressive. Identification of potential histological and pathological markers of aggressive breast tumors is already an area of active research, and one that should receive further funding. In addition, while it may not be considered ethical to randomize women with the small, node-negative tumors that would be identified by the technologies of focus in this review to receive active treatment or conservative management, it may be feasible to conduct retrospective observational studies to compare long-term outcomes among women of similar demographic and clinical profiles who had their cancers identified through more vs. less intense screening regimens to better quantify the benefits of early detection.

References

Alexander FE. The Edinburgh Randomized Trial of Breast Cancer Screening. Journal of the National Cancer Institute. Monographs. 1997(22):31-35.

American College of Radiology. ACR Maintains its Coding Recommendation for Breast Tomosynthesis. Accessed at: http://www.acr.org/News-Publications/News/News-Articles/2013/Economics/20131114-ACR-Maintains-Its-Coding-Recommendation-for-Breast-Tomosynthesis, December 2013.

American College of Radiology. CMS Establishes Breast Tomosynthesis Values in 2015 MPFS Final Rule. Accessed at: http://www.acr.org/News-Publications/News/News-Articles/2014/Economics/20141105-CMS-Establishes-Values-for-Breast-Tomosynthesis-in-2015-Final-Rule, November 2014.

American College of Radiology: Breast imaging reporting and data system (BI-RADS). 4th ed. Reston, VA: American College of Radiology; 2003.

An act to add and repeal Section 123222.3 of the Health and Safety Code, relating to mammograms. In: Legislature CS, ed. SB1538.

Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. Bmj. Oct 15 1988;297(6654):943-948.

Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. Journal of the National Cancer Institute. Monographs. 1997(22):63-67.

Arleo EK, Saleh M, Ionescu D, Drotman M, Min RJ, Hentel K. Recall rate of screening ultrasound with automated breast volumetric scanning (ABVS) in women with dense breasts: a first quarter experience. Clin Imaging. 2014 Jul-Aug;38(4):439-444.

Armstrong K, Moye E, Williams S, Berlin JA, Reynolds EE. Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians. Annals of internal medicine. Apr 3 2007;146(7):516-526.

Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. Journal of the National Cancer Institute. Sep 6 2006;98(17):1204-1214.

Barton MB, Moore S, Polk S, Shtatland E, Elmore JG, Fletcher SW. Increased patient concern after false-positive mammograms: clinician documentation and subsequent ambulatory visits. J Gen Intern Med. Mar 2001;16(3):150-156.

Barton MB, Morley DS, Moore S, et al. Decreasing women's anxieties after abnormal mammograms: a controlled trial. Journal of the National Cancer Institute. Apr 7 2004;96(7):529-538.

Berg WA, Blume JD, Adams AM, et al. Reasons women at elevated risk of breast cancer refuse breast MR imaging screening: ACRIN 6666. Radiology. Jan 2010;254(1):79-87.

Berg WA, Blume JD, Cormack JB, et al. Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. JAMA : the journal of the American Medical Association. May 14 2008;299(18):2151-2163.

Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. Radiology. Dec 2004;233(3):830-849.

Berg WA, Zhang Z, Lehrer D, 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. Apr 4 2012;307(13):1394-1404.

Berrington de Gonzalez A, Reeves G. Mammographic screening before age 50 years in the UK: comparison of the radiation risks with the mortality benefits. British journal of cancer. Sep 5 2005;93(5):590-596.

Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. The New England journal of medicine. Oct 27 2005;353(17):1784-1792.

Bevers TB, Anderson BO, Bonaccio E, et al. NCCN clinical practice guidelines in oncology: breast cancer screening and diagnosis. Journal of the National Comprehensive Cancer Network : JNCCN. Nov 2009;7(10):1060-1096.

Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg Breast Cancer Screening Trial: preliminary results on breast cancer mortality for women aged 39-49. Journal of the National Cancer Institute. Monographs. 1997a(22):53-55.

Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization. Cancer. Dec 1 1997b;80(11):2091-2099.

Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. The New England journal of medicine. Nov 22 2012;367(21):1998-2005.

Boice JD, Jr. Radiation and breast carcinogenesis. Medical and pediatric oncology. May 2001;36(5):508-513.

Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. The New England journal of medicine. Jan 18 2007;356(3):227-236.

Brancato B, Bonardi R, Catarzi S, et al. Negligible advantages and excess costs of routine addition of breast ultrasonography to mammography in dense breasts. Tumori. Nov-Dec 2007;93(6):562-566.

Breast-cancer screening with mammography in women aged 40-49 years. Swedish Cancer Society and the Swedish National Board of Health and Welfare. International journal of cancer. Journal international du cancer. Dec 11 1996;68(6):693-699.

Brem RF, Tabár L, Duffy SW, 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. 2014 Oct 17:132832. [Epub ahead of print]

Brett J, Bankhead C, Henderson B, Watson E, Austoker J. The psychological impact of mammographic screening. A systematic review. Psychooncology. Nov 2005;14(11):917-938.

Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. Annals of internal medicine. Apr 3 2007;146(7):502-510.

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. Aug 2000;21(4):325-336.

Carmichael A, Sami AS, Dixon JM. Breast cancer risk among the survivors of atomic bomb and patients exposed to therapeutic ionising radiation. European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology. Jun 2003;29(5):475-479.

Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Annals of internal medicine. Feb 4 2003;138(3):168-175.

Centers for Medicare and Medicaid Services CfMaMSMPFSL-UT, 2013. Accessed at: http://www.cms.gov/apps/physician-fee-schedule/search/search-criteria.aspx, August 2013.

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 Sep;32(9):1573-1578.

Chen J, Pee D, Ayyagari R, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. Journal of the National Cancer Institute. Sep 6 2006;98(17):1215-1226.

Christiansen CL, Wang F, Barton MB, et al. Predicting the cumulative risk of false-positive mammograms. Journal of the National Cancer Institute. Oct 18 2000;92(20):1657-1666.

Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. The lancet oncology. Jun 2013;14(7):583-589.

Conn. Gen. Stat. § 38a-503 a-M, 2009).

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

Corsetti V, Houssami N, Ferrari A, et al. Breast screening with ultrasound in women with mammography-negative dense breasts: evidence on incremental cancer detection and false positives, and associated cost. European journal of cancer. Mar 2008;44(4):539-544.

Corsetti V, Houssami N, Ghirardi M, et al. Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. European journal of cancer. May 2011;47(7):1021-1026.

Costantino JP, Gail MH, Pee D, et al. Validation studies for models projecting the risk of invasive and total breast cancer incidence. Journal of the National Cancer Institute. Sep 15 1999;91(18):1541-1548.

Cox B. Variation in the effectiveness of breast screening by year of follow-up. Journal of the National Cancer Institute. Monographs. 1997(22):69-72.

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

Cummings SR, Tice JA, Bauer S, et al. Prevention of breast cancer in postmenopausal women: approaches to estimating and reducing risk. Journal of the National Cancer Institute. Mar 18 2009;101(6):384-398.

D. K. Being less dense about breast cancer screening.

De Felice C, Savelli S, Angeletti M, et al. Diagnostic utility of combined ultrasonography and mammography in the evaluation of women with mammographically dense breasts. Journal of ultrasound. Sep 2007;10(3):143-151.

de Gelder R, Draisma G, Heijnsdijk EA, de Koning HJ. Population-based mammography screening below age 50: balancing radiation-induced vs prevented breast cancer deaths. British journal of cancer. Mar 29 2011;104(7):1214-1220.

de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiologic reviews. Jul 2011;33(1):111-121.

Department of Health and Human Services Maine Center for disease control and prevention. LD 1886 Resolve Work Group final report. 2012.

DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. CA: a cancer journal for clinicians. Nov-Dec 2011;61(6):409-418.

Destounis S, Arieno A, Morgan R. Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. J Clin Imaging Sci. 2014 Feb 25;4:9.

Drukteinis JS, Mooney BP, Flowers CI, Gatenby RA. Beyond mammography: new frontiers in breast cancer screening. The American journal of medicine. Jun 2013;126(6):472-479.

Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. The New England journal of medicine. Apr 16 1998;338(16):1089-1096.

Elmore JG, Fletcher SW. Overdiagnosis in breast cancer screening: time to tackle an underappreciated harm. Annals of internal medicine. Apr 3 2012;156(7):536-537.

Elwood JM, Cox B, Richardson AK. The effectiveness of breast cancer screening by mammography in younger women. Online J Curr Clin Trials. Feb 25 1993;Doc No 32:[23,227 words; 195 paragraphs].

Federal Register, The daily journal of The United States Government. Mammography Quality and Standards Act; Regulatory Amendments. 2014.

Feig SA, Yaffe MJ. Digital mammography. Radiographics. Jul-Aug 1998;18(4):893-901.

Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. Journal of the National Cancer Institute. 1998;90(18):1371-1388.

Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA. 2014 Jun 25;311(24):2499-1507.

Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years--update of mortality results in the Stockholm mammographic screening trial. Breast cancer research and treatment. Sep 1997b;45(3):263-270.

Frisell J, Lidbrink E. The Stockholm Mammographic Screening Trial: Risks and benefits in age group 40-49 years. Journal of the National Cancer Institute. Monographs. 1997a(22):49-51.

Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. Journal of the National Cancer Institute. Dec 20 1989;81(24):1879-1886.

Gail MH, Costantino JP, Bryant J, et al. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. Journal of the National Cancer Institute. Nov 3 1999;91(21):1829-1846.

Gartlehner G, Thaler KJ, Chapman A, et al. Adjunct ultrasonography for breast cancer screening in women at average risk: a systematic review. International journal of evidence-based healthcare. Jun 2013;11(2):87-93.

GE Healthcare U-S, a GE Healthcare Company. Current Coding Practices for Certain Breast Ultrasound Services. Accessed at: http://u-systems.com/healthcare/current-coding-practices-certain-breast-ultrasound-services, August 2013.

Girardi V, Tonegutti M, Ciatto S, Bonetti F. Breast ultrasound in 22,131 asymptomatic women with negative mammography. Breast. Apr 1 2013.

Giuliano V, Giuliano C. Improved breast cancer detection in asymptomatic women using 3D-automated breast ultrasound in mammographically dense breasts. Clin Imaging. May-Jun 2013;37(3):480-486.

Glasziou PP, Woodward AJ, Mahon CM. Mammographic screening trials for women aged under 50. A quality assessment and meta-analysis. The Medical journal of Australia. Jun 19 1995;162(12):625-629.

Glasziou PP. Meta-analysis adjusting for compliance: the example of screening for breast cancer. Journal of clinical epidemiology. Nov 1992;45(11):1251-1256.

Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. The Cochrane database of systematic reviews. 2013;6:CD001877.

Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev. 2006(4):CD001877.

Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet. Jan 8 2000;355(9198):129-134.

Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. AJR Am J Roentgenol. 2014 Sep;203(3):687-693.

H.R. 1302. Breast Density and Mammography Reporting Act of 2011. U.S. Congress. 2011.

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. Jul 30 2013.

Hagen AI, Kvistad KA, Maehle L, et al. Sensitivity of MRI versus conventional screening in the diagnosis of BRCA-associated breast cancer in a national prospective series. Breast. Aug 2007;16(4):367-374.

Hardesty LA, Kreidler SM, Glueck DH. Digital breast tomosynthesis utilization in the United States: a survey of physician members of the Society of Breast Imaging. J Am Coll Radiol. 2014 Jun;11(6):594-599.

Hartman AR, Daniel BL, Kurian AW, et al. Breast magnetic resonance image screening and ductal lavage in women at high genetic risk for breast carcinoma. Cancer. Feb 1 2004;100(3):479-489.

Hendrick RE, Pisano ED, Averbukh A, et al. Comparison of acquisition parameters and breast dose in digital mammography and screen-film mammography in the American College of Radiology Imaging Network digital mammographic imaging screening trial. AJR. American journal of roentgenology. Feb 2010;194(2):362-369.

Hendrick RE, Smith RA, Rutledge JH, 3rd, Smart CR. Benefit of screening mammography in women aged 40-49: a new meta-analysis of randomized controlled trials. Journal of the National Cancer Institute. Monographs. 1997(22):87-92.

Hofvind S, Thoresen S, Tretli S. The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer. Oct 1 2004;101(7):1501-1507.

Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM. Comparing screening mammography for early breast cancer detection in Vermont and Norway. Journal of the National Cancer Institute. Aug 6 2008;100(15):1082-1091.

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. Oct 2012;265(1):59-69.

Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. Breast. Apr 2013;22(2):101-108.

Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading--evidence to guide future screening strategies. Eur J Cancer. 2014 Jul;50(10):1799-1807

Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. Annals of internal medicine. Oct 18 2011;155(8):481-492.

Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Annals of internal medicine. Sep 3 2002;137(5 Part 1):347-360.

Ind. Code § 25-22.5-13.2. 2013.

Independent UKPoBCS. The benefits and harms of breast cancer screening: an independent review. Lancet. Nov 17 2012;380(9855):1778-1786.

Jensen A, Geller BM, Gard CC, et al. Performance of diagnostic mammography differs in the United States and Denmark. International journal of cancer. Journal international du cancer. Oct 15 2010;127(8):1905-1912.

Jorgensen KJ, Gotzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. Bmj. 2009;339:b2587.

Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Annals of internal medicine. Apr 3 2012;156(7):491-499.

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

Keen JD, Keen JE. What is the point: will screening mammography save my life? BMC Med Inform Decis Mak. 2009;9:18.

Kelly KM, Dean J, Comulada WS, Lee SJ. Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. Eur Radiol. Mar 2010;20(3):734-742.

Kelly KM, Dean J, Lee SJ, Comulada WS. Breast cancer detection: radiologists' performance using mammography with and without automated whole-breast ultrasound. Eur Radiol. Nov 2010;20(11):2557-2564.

Kelly KM, Richwald GA. Automated Whole-Breast Ultrasound: Advancing the Performance of Breast Cancer Screening. Seminars in Ultrasound, CT and MRI. 2011;32(4):273-280.

Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA : the journal of the American Medical Association. Jul 3 1996;276(1):33-38.

Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA : the journal of the American Medical Association. Jan 11 1995;273(2):149-154.

Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Annals of internal medicine. Oct 18 2011;155(8):493-502.

Kerlikowske K, Zhu W, Hubbard RA, et al. Outcomes of screening mammography by frequency, breast density, and postmenopausal hormone therapy. JAMA internal medicine. May 13 2013;173(9):807-816.

Kerlikowske K. Efficacy of screening mammography among women aged 40 to 49 years and 50 to 69 years: comparison of relative and absolute benefit. Journal of the National Cancer Institute. Monographs. 1997(22):79-86.

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

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

Kopans DB, Skaane P. Digital breast tomosynthesis: a better mammogram. Radiology. Jun 2013;267(3):968-969.

Korpraphong P, Limsuwarn P, Tangcharoensathien W, Ansusingha T, Thephamongkhol K, Chuthapisith S. Improving breast cancer detection using ultrasonography in asymptomatic women with non-fatty breast density. Acta Radiol. 2014 Oct;55(8):903-908.

Kriege M, Brekelmans CT, Boetes C, et al. Differences between first and subsequent rounds of the MRISC breast cancer screening program for women with a familial or genetic predisposition. Cancer. Jun 1 2006;106(11):2318-2326.

Kriege M, Brekelmans CT, Boetes C, et al. Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. The New England journal of medicine. Jul 29 2004;351(5):427-437.

Kuhl CK, Schmutzler RK, Leutner CC, et al. Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. Radiology. Apr 2000;215(1):267-279.

Kuhl CK, Schrading S, Leutner CC, et al. Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. Nov 20 2005;23(33):8469-8476.

Kuhl C, Weigel S, Schrading S, et al. Prospective multicenter cohort study to refine management recommendations for women at elevated familial risk of breast cancer: the EVA trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. Mar 20 2010;28(9):1450-1457.

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 Aug 1;32(22):2304-2310.

Leach MO, Boggis CR, Dixon AK, et al. Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). Lancet. May 21-27 2005;365(9473):1769-1778.

Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. AJR. American journal of roentgenology. Jun 2003;180(6):1675-1679.

Lee CH, Dershaw DD, Kopans D, et al. Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. Journal of the American College of Radiology : JACR. Jan 2010;7(1):18-27.

Lee CI, Cevik M, Alagoz O, et al. Comparative effectiveness of combined digital mammography and tomosynthesis screening for women with dense breasts. Radiology. 2014 Oct 13:141237. [Epub ahead of print].

Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. Journal of clinical epidemiology. Jan 2009;62(1):5-12.

Lehman CD, Blume JD, Weatherall P, et al. Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. Cancer. May 1 2005;103(9):1898-1905.

Lehman CD, Isaacs C, Schnall MD, et al. Cancer yield of mammography, MR, and US in high-risk women: prospective multi-institution breast cancer screening study. Radiology. Aug 2007;244(2):381-388.

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. Annals of the Academy of Medicine, Singapore. Oct 2012;41(10):432-439.

Library of Congress. S.2622 - Breast Density and Mammography Reporting Act of 2014. https://www.congress.gov/bill/113th-congress/senate-bill/2622/related-bills

Linver MN, Osuch JR, Brenner RJ, Smith RA. The mammography audit: a primer for the mammography quality standards act (MQSA). AJR Am J Roentgenol. Jul 1995;165(1):19-25.

Lipkus IM, Halabi S, Strigo TS, Rimer BK. The impact of abnormal mammograms on psychosocial outcomes and subsequent screening. Psychooncology. Sep-Oct 2000;9(5):402-410.

Lord SJ, Lei W, Craft P, 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. European journal of cancer. Sep 2007;43(13):1905-1917.

Lourenco AP, Barry-Brooks M, Baird G, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. Radiology. 2014 Sep 22:140317. [Epub ahead of print]

Maestro C, Cazenave F, Marcy PY, Bruneton JN, Chauvel C, Bleuse A. Systematic ultrasonography in asymptomatic dense breasts. Eur J Radiol. Feb 1998;26(3):254-256.

Mahoney MC, Newell MS. Screening MR Imaging Versus Screening Ultrasound. Pros and Cons. Magnetic Resonance Imaging Clinics of North America. 2013.

Malone KE, Daling JR, Doody DR, et al. Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years. Cancer research. Aug 15 2006;66(16):8297-8308.

Mann RM, Kuhl CK, Kinkel K, Boetes C. Breast MRI: guidelines from the European Society of Breast Imaging. European radiology. Jul 2008;18(7):1307-1318.

McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. J Natl Cancer Inst. 2014 Oct 13;106(11).

Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. Cmaj. Nov 15 1992a;147(10):1477-1488.

Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. Cmaj. Nov 15 1992b;147(10):1459-1476.

Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. Journal of the National Cancer Institute. Sep 20 2000;92(18):1490-1499.

Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study: update on breast cancer mortality. Journal of the National Cancer Institute. Monographs. 1997(22):37-41.

Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. Annals of internal medicine. Sep 3 2002;137(5 Part 1):305-312.

Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. Cancer causes & control : CCC. Feb 2010;21(2):275-282.

Morris EA, Liberman L, Ballon DJ, et al. MRI of occult breast carcinoma in a high-risk population. AJR Am J Roentgenol. Sep 2003;181(3):619-626.

Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet. Dec 9 2006;368(9552):2053-2060.

National Comprehensive Cancer Network. Breast cancer screening and diagnosis. NCCN Clinical Practice Guidelines in Oncology. 2013(1.2013).

National Comprehensive Cancer Network. Breast cancer screening and diagnosis. NCCN Clinical Practice Guidelines in Oncology. 2013(1.2013).

Ng AK, Travis LB. Radiation therapy and breast cancer risk. Journal of the National Comprehensive Cancer Network : JNCCN. Nov 2009;7(10):1121-1128.

NHSBSP- 2012 NHSBSP, National Collation of Breast Interval Data: Screening years 1st April 2003 - 31st March 2005 (2012). Accessed at: http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp-occasional-report1203.pdf, August 2013. .

Nothacker M, Duda V, Hahn M, 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.

Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet. Mar 16 2002;359(9310):909-919.

Nystrom L, Rutqvist LE, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet. Apr 17 1993;341(8851):973-978.

O'Donoghue C, Eklund M, Ozanne EM, Esserman LJ. Aggregate cost of mammography screening in the United States: comparison of current practice and advocated guidelines. Ann Intern Med. 2014 Feb 4;160(3):145. doi: 10.7326/M13-1217.

Olivotto IA, Kan L, Coldman AJ. False positive rate of screening mammography. The New England journal of medicine. Aug 20 1998;339(8):560; author reply 563.

Parris T, Wakefield D, Frimmer H. Real world performance of screening breast ultrasound following enactment of Connecticut Bill 458. Breast J. Jan-Feb 2013;19(1):64-70.

Partin MR, Casey-Paal AL, Slater JS, Korn JE. Measuring mammography compliance: lessons learned from a survival analysis of screening behavior. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. Aug 1998;7(8):681-687.

Pisano ED, Cole EB, Hemminger BM, et al. Image processing algorithms for digital mammography: a pictorial essay. Radiographics. Sep-Oct 2000;20(5):1479-1491.

Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. The New England journal of medicine. Oct 27 2005;353(17):1773-1783.

Pisano ED, Gatsonis CA, Yaffe MJ, et al. American College of Radiology Imaging Network digital mammographic imaging screening trial: objectives and methodology. Radiology. Aug 2005;236(2):404-412.

Pisano ED, Hendrick RE, Yaffe MJ, et al. Diagnostic accuracy of digital versus film mammography: exploratory analysis of selected population subgroups in DMIST. Radiology. Feb 2008;246(2):376-383.

Pisano ED, Zong S, Hemminger BM, et al. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. J Digit Imaging. Nov 1998;11(4):193-200.

Pisano ED, Yaffe MJ. Breast cancer screening: should tomosynthesis replace digital mammography? JAMA. 2014 Jun 25;311(24):2488-2489.

Podo F, Sardanelli F, Canese R, 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. Journal of experimental & clinical cancer research : CR. Sep 2002;21(3 Suppl):115-124.

Port ER, Park A, Borgen PI, Morris E, Montgomery LL. Results of MRI screening for breast cancer in high-risk patients with LCIS and atypical hyperplasia. Annals of surgical oncology. Mar 2007;14(3):1051-1057.

Quality mammography standards--FDA. Final rule. Fed Regist. Oct 28 1997;62(208):55852-55994.

Ries LAG, Melbert D, Krapcho M, et al., eds. SEER Cancer Statistics Review, 1975-2005. Bethesda, MD: National Cancer Institute; 2008.

Ringash J. Preventive health care, 2001 update: screening mammography among women aged 40-49 years at average risk of breast cancer. Cmaj. Feb 20 2001;164(4):469-476.

Rockhill B, Spiegelman D, Byrne C, Hunter DJ, Colditz GA. Validation of the Gail et al. model of breast cancer risk prediction and implications for chemoprevention. Journal of the National Cancer Institute. Mar 7 2001;93(5):358-366.

Ronckers CM, Erdmann CA, Land CE. Radiation and breast cancer: a review of current evidence. Breast Cancer Res. 2005;7(1):21-32.

Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R, Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. AJR Am J Roentgenol. Jun 2013;200(6):1401-1408.

Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. Radiology. Oct 2006;241(1):55-66.

S.S.B. 6050. An act relating to communication of mammographic breast density information to patients. Washington State Legislature. 2014. http://apps.leg.wa.gov/documents/billdocs/2013-14/Pdf/Bills/Senate%20Bills/6050-S.pdf

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

Saslow D, Boetes C, Burke W, et al. American cancer society guidelines for breast screening with MRI as an adjunct to mammography. CA: a cancer journal for clinicians. Mar-Apr 2007;57(2):75-89.

Scaf-Klomp W, Sanderman R, van de Wiel HB, Otter R, van den Heuvel WJ. Distressed or relieved? Psychological side effects of breast cancer screening in The Netherlands. J Epidemiol Community Health. Dec 1997;51(6):705-710.

Scheel JR, Lee JM, Sprague BL, Lee CI, Lehman CD. Screening ultrasound as an adjunct to mammography in women with mammographically dense breasts [published online ahead of print June 21, 2014]. Am J Obstet Gynecol. doi: 10.1016/j.ajog.2014.

Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Annals of internal medicine. Nov 17 2009;151(10):716-726, W-236.

Sechopoulos I. A review of breast tomosynthesis. Part I. The image acquisition process. Med Phys. Jan 2013;40(1):014301.

SEER Cancer Statistics 2010: National Cancer Institute. Surveillance Epidemiology and End Results Database. SEER Cancer Statistics 2010. Accessed at: http://seer.cancer.gov/statfacts/html/breast.html, August 2013.

Senate Bill Report S.S.B. 6050. An act relating to communication of mammographic breast density information to patients. Washington State Legislature. 2014. http://apps.leg.wa.gov/documents/billdocs/2013-14/Pdf/Bill%20Reports/Senate/6050-S%20SBR%20APS%2014.pdf

Shapiro S, Venet W, Strax P, Venet L. Current results of the breast cancer screening randomized trial: the health insurance plan (HIP) of greater New York study. In: Miller NEDaAB, ed. Screening for breast cancer. Toronto Lewiston, N.Y.: Huber; 1988:3-15.

Shapiro S. Periodic screening for breast cancer: the HIP Randomized Controlled Trial. Health Insurance Plan. Journal of the National Cancer Institute. Monographs. 1997(22):27-30.

Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: a cancer journal for clinicians. Jan 2013;63(1):11-30.

Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology. Apr 2013;267(1):47-56.

Skaane P, Bandos AI, Gullien R, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. Eur Radiol. Apr 4 2013.

Skaane P, Bandos AI, Eben EB, 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 Jun;271(3):655-663.

Smart CR, Hendrick RE, Rutledge JH, 3rd, Smith RA. Benefit of mammography screening in women ages 40 to 49 years. Current evidence from randomized controlled trials. Cancer. Apr 1 1995;75(7):1619-1626.

Smith RA, Brooks D, Cokkinides V, Saslow D, Brawley OW. Cancer screening in the United States, 2013: a review of current American Cancer Society guidelines, current issues in cancer screening, and new

guidance on cervical cancer screening and lung cancer screening. CA: a cancer journal for clinicians. Mar-Apr 2013;63(2):88-105.

Smith-Bindman R, Chu PW, Miglioretti DL, et al. Comparison of screening mammography in the United States and the United kingdom. JAMA : the journal of the American Medical Association. Oct 22 2003;290(16):2129-2137.

Solis Womens Health. Henda's Law; Supplemental screening for women with dense breast tissue and increased risk. .

South Jersey Radiology. 3D Digital Tomosynthesis Mammography. Accessed at: http://www.sjra.com/services-list/3d-digital-tomosynthesis-mammography/, August 2013.

State of New Jersey 215th Legislature. Fiscal Note- Senate Bill No. 729. 2013. http://www.njleg.state.nj.us/2012/Bills/S1,000/792_F5.PDF

Stoblen F, Landt S, Stelkens-Gebhardt R, Sehouli J, Rezai M, Kummel S. First evaluation of the diagnostic accuracy of an automated 3D ultrasound system in a breast screening setting. Anticancer research. Aug 2011;31(8):2569-2574.

Stout NK, Lee SJ, Schechter CB, et al. Benefits, harms, and costs for breast cancer screening after US implementation of digital mammography [published online ahead of print May 28, 2014. J Natl Cancer Inst. doi:10.1093/jnci/dju092.

Stoutjesdijk MJ, Boetes C, Jager GJ, et al. Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. Journal of the National Cancer Institute. Jul 18 2001;93(14):1095-1102.

Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer. May 15 1995;75(10):2507-2517.

Tabar L, Fagerberg G, Duffy SW, Day NE. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. J Epidemiol Community Health. Jun 1989;43(2):107-114.

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

The benefits and harms of breast cancer screening: an independent review. Lancet. Nov 17 2012;380(9855):1778-1786.

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. Annals of internal medicine. Mar 4 2008;148(5):337-347.

Tilanus-Linthorst MM, Obdeijn IM, Bartels KC, de Koning HJ, Oudkerk M. First experiences in screening women at high risk for breast cancer with MR imaging. Breast cancer research and treatment. Sep 2000;63(1):53-60.

Titus-Ernstoff L, Tosteson AN, Kasales C, et al. Breast cancer risk factors in relation to breast density (United States). Cancer causes & control : CCC. Dec 2006;17(10):1281-1290.

Tosteson AN, Stout NK, Fryback DG, et al. Cost-effectiveness of digital mammography breast cancer screening. Annals of internal medicine. Jan 1 2008;148(1):1-10.

Trecate G, Vergnaghi D, Manoukian S, et al. MRI in the early detection of breast cancer in women with high genetic risk. Tumori. Nov-Dec 2006;92(6):517-523.

U.S. Census Bureau, 2010: U.S. Census Bureau. New England Demographic Profile, 2010. Accessed at: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?fpt=table, October 2013.

United States Food and Drug Administration. Mammography Quality Standards Act and Program. 2013; http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/FacilityScorecard/ucm113858.htm. Accessed July 15, 2013, 2013.

University of Virginia Health System UoVHSTDMAahucsb-cba--. Accessed at: http://uvahealth.com/services/breast-care/breastscreening/atproxiedcontent.2012-09-28.9643159495, August 2013.

U.S. Food and Drug Administration. Affirm Breast Biopsy Guidance System 510 (k) Summary. Accessed at: http://www.accessdata.fda.gov/cdrh_docs/pdf12/k122836.pdf, November 2014.

USPSTF, U.S. Preventive Services Task Force, Chemoprevention for Breast Cancer: Recommendations and Rationale (2002). Accessed at: http://www.uspreventiveservicestaskforce.org/3rduspstf/breastchemo/breastchemorr.pdf, August 2013.

Venturini E, Losio C, Panizza P, 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. Aug 2013;268(2):347-355.

Viala J, Gignier P, Perret B, et al. Stereotactic vacuum-assisted biopsies on a digital breast 3D-tomosynthesis system. Breast J. Jan-Feb 2013;19(1):4-9.

Vigeland E, Klaasen H, Klingen TA, Hofvind S, Skaane P. Full-field digital mammography compared to screen film mammography in the prevalent round of a population-based screening programme: the Vestfold County Study. European radiology. Jan 2008;18(1):183-191.

Visvanathan K, Hurley P, Bantug E, et al. Use of Pharmacologic Interventions for Breast Cancer Risk Reduction: American Society of Clinical Oncology Clinical Practice Guideline. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. Jul 8 2013.

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. Annals of internal medicine. May 6 2008;148(9):671-679.

Warner E, Plewes DB, Hill KA, et al. Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. JAMA : the journal of the American Medical Association. Sep 15 2004;292(11):1317-1325.

Warner E, Plewes DB, Shumak RS, et al. Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. Aug 1 2001;19(15):3524-3531.

Washington Radiology Associates. 3D Mammography. Accessed at: http://www.washingtonradiology.com/bca/, August 2013.

Washington State Legislature. 2014. S.B. 6050: History of the bill. Concerning communication of mammographic breast density information to patients. http://apps.leg.wa.gov/billinfo/summary.aspx?bill=6050

Washington State Medical Association. How did we do?: WSMA 2014 legislative summary. 2014.

Weigert J, Steenbergen S. The connecticut experiment: the role of ultrasound in the screening of women with dense breasts. Breast J. Nov-Dec 2012;18(6):517-522.

Weil JG, Hawker JI. Positive findings of mammography may lead to suicide. Bmj. Mar 8 1997;314(7082):754-755.

Welch HG, Black WC. Overdiagnosis in cancer. Journal of the National Cancer Institute. May 5 2010;102(9):605-613.

Welch HG. Overdiagnosis and mammography screening. Bmj. 2009;339:b1425.

Whittemore AS, Gong G, John EM, et al. Prevalence of BRCA1 mutation carriers among U.S. non-Hispanic Whites. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. Dec 2004;13(12):2078-2083.

Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer. May 1976;37(5):2486-2492.

Woloshin S, Schwartz LM, Welch HG. The risk of death by age, sex, and smoking status in the United States: putting health risks in context. Journal of the National Cancer Institute. Jun 18 2008;100(12):845-853.

Woodward V, Webb C. Women's anxieties surrounding breast disorders: a systematic review of the literature. J Adv Nurs. Jan 2001;33(1):29-41.

Wu D, Perez A. A limited Review of Over Diagnosis Methods and Long Term Effects in Breast Cancer Screening. Oncol Rev. Sep 1 2011;5(3):143-147.

Yaffe MJ, Mainprize JG. Risk of radiation-induced breast cancer from mammographic screening. Radiology. Jan 2011;258(1):98-105.

Yaghjyan L, Colditz GA, Rosner B, Tamimi RM. Mammographic breast density and subsequent risk of breast cancer in postmenopausal women according to the time since the mammogram. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. Jun 2013;22(6):1110-1117.

Yang WT. Emerging Techniques and Molecular Imaging in Breast Cancer. Seminars in Ultrasound, CT and MRI. 2011;32(4):288-299.

Yu J, Park A, Morris E, Liberman L, Borgen PI, King TA. MRI screening in a clinic population with a family history of breast cancer. Annals of surgical oncology. Feb 2008;15(2):452-461.

Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: follow-up study. Bmj. Mar 25 2006;332(7543):689-692.

Zahl PH, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Archives of internal medicine. Nov 24 2008;168(21):2311-2316.

Zuley ML, Guo B, Catullo VJ, et al. Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. Radiology. 2014 Jun;271(3):664-671.

Appendix A

Search Strategies

Search A

Scope: Use of digital breast tomosynthesis versus digital mammography as a frontline screening tool.

Ovid Search Terms:
1. (breast and screen* and mammogra* and (3D or 3 dimension* or three dimension*)).ab,kw,ti.
2. exp Breast Neoplasms/ [Diagnosis, Radiography]
3. exp Mass Screening/
4. screen*.mp
5. 3 or 4
6. 2 and 5
7. Imaging, Three-Dimensional/
8. (tomosyn* or 3D or 3 dimension* or three dimension*).ti,ab,kw.
9. 7 or 8
10. 6 and 9
11. 1 or 10
12. limit 11 to (english language and humans and yr="2008-current")

Embase Search Terms:
1. 'breast tumor'/exp/mj OR 'breast tumor'/exp/dm_di OR 'breast'/exp
2. screen*
3. mammogra*
4. 'tomography'/de AND (tomosyn* OR '3d' or 3 NEXT/1 dimension* OR three NEXT/1 dimension*)
5. #1 AND #2 AND #2 AND #4
6. mammogra* AND tomosyn*
7. breast:ti AND (mammogra*:ab,ti OR echomammography) AND (tomosyn*:ti OR 3d:ti OR (3 NEXT/1 dimension*):ti OR (three NEXT/1 dimension*):ti)
8. #5 OR #6 OR #7
9. #8 NOT ('case report'/exp OR letter/it OR ('review'/it OR 'short survey'/it NOT (random* OR 'systematic review' OR metaanalysis or 'meta analysis')) OR 'conference abstract'/it) NOT ([animals]/lim NOT [humans]/lim) AND [english]/lim
10. #9 AND ([priority journals]/lim AND [2008-2014]/py

Include: Study populations comprising of asymptomatic middle-aged women (40-74) at average risk for breast cancer undergoing routine screening with digital mammography or DBT/digital mammography.
Sources: Health technology assessments, systematic reviews, RCTs, comparative studies, case control studies, case series

Search B

Scope: Use of automated and handheld ultrasound as well as magnetic resonance imaging for supplemental screening in women with dense breast tissue.

Ovid Search Terms:
1. (breast and screen* and mammogra* and (magnetic resonance or ultraso* or sonogra* or 'image enhancement' or MRI)).ab,kw,ti.
2. exp Breast Neoplasms/ [Diagnosis, Radiography]
3. exp Mass Screening/
4. screen*.mp
5. 3 or 4
6. 2 and 5
7. Magnetic Resonance Imaging/
8. exp Radiography Image Enhancement/
9. exp Ultrasonography, Mammary/
10. Ultrasonography/
11. 7 or 8 or 9 or 10
12. (magnetic resonance or MRI or ultraso* or sonogra* or 'image enhancement').mp
13. 6 and 11 and 12
14. 1 or 13
15. limit 14 to (english language and humans and yr="2013-current")

Embase Search Terms:
1. 'breast tumor'/exp/mj OR 'breast tumor'/exp/dm_di OR ('breast'/exp AND (dense OR density OR densities)) OR breast NEAR/3 (dense OR density OR densities)
2. screen*
3. mammogra*
4. 'nuclear magnetic resonance'/exo OR 'image enhancement'/de OR 'echomammography'/de OR 'ultrasound'/mj
5. #1 AND #2 AND #2 AND #4
6. breast:ti AND dens*:ab,ti AND (mammogra*:ab,ti OR echomammography) AND (handheld:ti OR 'hand held':ti OR imaging:ti OR mri:ti OR radiogra*:ti OR ultraso*:ti)
7. #5 OR #6
8. #7 NOT ('case report'/exp OR letter/it OR ('review'/it OR 'short survey'/it NOT (random* OR 'systematic review' OR metaanalysis or 'meta analysis')) OR 'conference abstract'/it) NOT ([animals]/lim NOT [humans]/lim) AND [english]/lim
9. #8 AND ([priority journals]/lim AND [2013-2014]/py

Include: Study populations comprising of asymptomatic middle-aged women (40-74) with dense breasts at average risk for breast cancer. Other requirements include: Women who have a "normal" result from digital mammography (as defined by BI-RADS categories 1 or 2), and; Supplemental screening with MRI, HHUS, or ABUS
Sources: Health technology assessments, systematic reviews, RCTs, comparative studies, case control studies, case series

Appendix B

Detailed Study Tables

Table A1: Description of the studies.

Study	Test	Years of Study	Population	N	Follow-up	Age (Years)
DBT						
Ciatto 2013 STORM Italy	Selenia Dimensions, Hologic	2011-2012	Asymptomatic screening D3, D4 subgroup	1,127 D3/D4 7,292 total	None	NR for D3/D4 subgroup Mean 58 Range 48-71
Destounis 2014 United States	Selenia Dimensions, Hologic SecurView, Hologic	2011	Asymptomatic screening Most women in D3, D4	524 DM 524 DBT	1-year follow-up	Average 59
Friedewald 2014 United States	Selenia Dimensions, Hologic	2010-2012	Asymptomatic screening	281,187 DM 173,663 DBT	None	Mean 57 for DM Mean 56 for DM + DBT
Greenberg 2014 United States	Selenia Dimensions, Hologic	2011-2012	Asymptomatic screening	38,674 DM 20,943 DBT	None	Mean 59.5 for DM Mean 59.6 for DM + DBT
Haas 2013 United States	Selenia Dimensions, Hologic	2011-2012	Asymptomatic screening	13,158 DM 6,100 DBT	None	NR D3/D4 subgroup Mean 56 Range <40 to >70
Lourenco 2014 United States	Selenia Dimensions, Hologic	2011-2013	Asymptomatic screening	12,577 DM 12,921 DBT	None	Mean 54.6 for DM Mean 55.3 for DBT
McCarthy 2014 United States	Selenia Dimensions, Hologic	2011-2013	Asymptomatic screening	10,728 DM 15,571 DBT	1-year follow-up but overlap with DBT study period	Mean 56.9 for DM Mean 56.7 for DBT

Study	Test	Years of Study	Population	N	Follow-up	Age (Years)
Rose 2013 United States	Selenia Dimensions, Hologic	2011-2012	Asymptomatic Screening Elected to have DBT D3, D4 subgroup	4,666 total	None	NR for subgroup
Skaane 2013 Norway	Selenia Dimensions, Hologic	2010-2011	Asymptomatic Screening	12,621 total	None	NR D3/D4 subgroup Mean 59 50-69
MRI						
Kuhl 2014 Germany	27 to 33 axial-weighted gradient echo images 2 radiologists	2009-2011	Asymptomatic screening "Dense" breasts	433 women, 606 MRI images	2-year validation period No cancers detected at second screening	Mean 54 Range 25-73
HHUS						
Maestro 1998 France	7.5, 10, or 13 Hz Esaote Biomedica Operator NR	1994-1995	Asymptomatic Screening "Dense" breasts	350	Variable	Mean 52
Buchberger 2000 Austria	5-12 MHz ATL Radiologist	1996-2000	Asymptomatic Screening D2, D3, D4	8,103	None	Mean 48 Range 35-78
Kaplan 2001 United States	7-12 MHz GE Technician	1998-2000	Asymptomatic Screening D3, D4 subgroup	1,862	Variable 72 followed for 1 year: 0 cancer	35-87
Kolb 2002 United States	5-12 MHz ATL 1 radiologist	1995-2000	Asymptomatic Screening D2, D3, D4	12,193 examinations in 4897 women	Variable. All participant with biopsy followed for 1 year	Mean 55

Study	Test	Years of Study	Population	N	Follow-up	Age (Years)
Crystal 2003 Israel	5-12 MHz ATL Radiologist	2000-2002	Asymptomatic Screening D2, D3, D4	1,517	None	Mean 52 Range 31-84
Leconte 2003 Belgium	4.8-9.6 MHz Elegra, Siemens Radiologist	2000-2001	Mix 3% symptomatic, 24% breast cancer follow-up, 76% screening D1-D4	4,236 3084 screening	None	NR
Brancato 2007 Italy	10-14 MHz Esaote Technos Radiologists	2003-2006	Asymptomatic Screening D3, D4 subgroup	5,227	None	NR for subgroup
De Felice 2007 Italy	10-13 MHz Aloka, GE Radiologist	2000-2006	Asymptomatic Screening D3, D4 subgroup	1,754	None	NR
Corsetti 2008 Italy	7.5-10 MHz Aloka Pro Sound Physician	2000-2007	Asymptomatic Screening D3, D4 subgroup	9,157	None	Mean 52
Korpraphong 2014 Thailand	10-14 MHz GE Radiologist	2006-2007	Asymptomatic screening D2, D3, and D4	14,483	~2 years	Mean 49.6
Corsetti 2011 Italy	7.5-10 MHz Aloka Pro Sound Physician	2001-2006	Asymptomatic Screening D3, D4 subgroup	7,224 examinations of 3356 women	One year	NR for subgroup
Hooley 2012 United States	12.5 – 17.5 MHz Phillips IU22 Technician	2009-2010	Asymptomatic Screening D3, D4 subgroup	935	One year	Mean 52 Range 29-89

Study	Test	Years of Study	Population	N	Follow-up	Age (Years)
Leong 2012 Singapore	7-10 MHz Toshiba PowerVision Technician	2002-2004	Asymptomatic Screening D3, D4 subgroup	141	One to two years	Mean 45 Range 30-64
Weigert 2012 United States	12.5 MHz Technician	2009-2010	Asymptomatic Screening D3, D4 subgroup	8,647	None	NR
Chae 2013 Korea	iU22, Philips Healthcare	2008-2009	Asymptomatic Screening Most women in D3, D4	20,864	Two years	Mean 52 Range 22-91
Girardi 2013 Italy	12 MHz Radiologist	2009-2010	Asymptomatic Screening D3, D4 subgroup	9,960	None	Overall 51, range 33-84 NR D3/D4
Parris 2013 United States	12 MHz Philips HDI 5000 Technician	2009-2010	Asymptomatic Screening D3, D4 subgroup	5,519	None	Mean 54
Venturini 2013 Italy	Technos US system, Esaote Biomedica 2 Radiologists	2010-2011	Asymptomatic Screening D3, D4 subgroup	1,666	3-6 months follow-up for BI- RADS category 3 findings only	Range 40-49
Berg 2012 ACRIN 6666 United States	≥ 12 MHz Radiologist	2004-2006	High-risk D3 in at least 1 quadrant of 1 breast	2,659	One year for each of 3 rounds	Median 55 Range 25-91

Study	Test	Years of Study	Population	N	Follow-up	Age (Years)
ABUS						
Arleo 2014	S2000, Siemens Healthcare	August-October 2013	Asymptomatic Screening All women in D3, D4	558	None	Mean 53
Kelly 2010 United States	SonoCine	2003-2007	Asymptomatic Screening D3, D4 subgroup	6,425 examinations for 4419 women	One year	Mean 53 Range 24 to 89
Stoblen 2011 Germany	SomoV, U-Systems	2008	Asymptomatic Screening Majority D2	304	None	Mean 58 Range 50 to 69
Brem 2014 United States	SomoV, U-Systems	2009-2011	Asymptomatic screening All women in D3, D4	15,318	1-year follow-up	Mean 53
Giuliano 2013 United States	SomoV, U-Systems	2010-2011	Asymptomatic Screening Wolf density ≥ 50%	3,418	One year	Mean 57 Range <50 to >70

Table A2: Test characteristics of MRI, HHUS, and ABUS.

Study	TP (N)	FP (N)	FN (N)	TN (N)	Sens (%)	Spec (%)	PPV1 (%)	NPV (%)	PPV3 (%)	Recall rate (per 1,000)	Biopsy rate (per 1,000)	Cancer detection rate (per 1,000)
MRI												
Kuhl 2014	11	NR	0	553	100	94.3	24.4	100	NR	NR	NR	18.2
HHUS												
Maestro 1998	2	46	0*	302	100	86.8	4.2	100	13.3	137.1	42.9	5.7
Buchberger 2000	32	241	0*	7830	100	97.0	11.7	100	9.9	33.7	39.9	3.9
Kaplan 2001	5	245	0*	1612	100	86.8	2.0	100	5.4	134.3	49.4	2.7
Kolb 2002	31	NR	NR	NR	-	-	-	-	10.6	-	23.9	2.5
Crystal 2003	7	90	0*	1420	100	94.0	7.2	100	18.4	63.9	25.0	4.6
Leconte 2003	11	NR	NR	NR	-	-	-	-	NR	NR	NR	-
Brancato 2007	2	106	0*	5119	100	98.0	1.9	100	3.2	20.7	11.9	0.4
De Felice 2007	12	175	0*	1567	100	90.0	6.4	100	6.4	106.6	106.6	6.4
Corsetti 2008	37	412	0*	8708	100	95.5	8.2	100	5.7	49.0	9.1	4.0
Corsetti 2011	32	395	8	6769	80.0	94.5	7.5	99.9	7.5	59.3	59.3	7.5
Hooley 2012	3	50	0	882	100	94.6	5.7	100	5.7	56.7	56.7	3.2
Leong 2012	2	22	0	117	100	84.2	8.3	100	12.5	170.2	113.5	14.2
Weigert 2012	28	401	1	8217	96.6	95.3	6.5	100	6.7	49.6	48.3	3.2
Chae 2013	24	194	0	NR	100	94.8	5.3	0	NR	55	25.8	2.9
Girardi 2013	22	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR	2.2
Parris 2013	10	175	0*	5334	100	96.8	5.4	100	5.5	33.5	32.8	1.8
Venturini 2013	14	73	NR	1579	-	95.6	2.3	-	8.3	52.2	11.9	6.0
Berg 2012 R1	14	423	2	1914	87.5	81.9	3.2	99.9	6.8	185.7	88.0	5.9
ACRIN 6666 R3	9	326	14	1934	39.1	91.8	5.0	99.3	NR	85.0	NR	4.2

Study	TP (N)	FP (N)	FN (N)	TN (N)	Sens (%)	Spec (%)	PPV1 (%)	NPV (%)	PPV3 (%)	Recall rate (per 1,000)	Biopsy rate (per 1,000)	Cancer detection rate (per 1,000)
ABUS												
Arelo 2014	0	98	NR	453	NR	82.2	0	100	0	188.2	19.7	0
Kelly 2010	23	442	10	5657	69.7	92.8	4.9	99.8	30.7	75.8	12.2	3.8
Brem 2014	112	4252	0	10954	100	72.0	2.6	100	9.8	284.9	74.3	7.3
Stoblen 2011	0	60	0*	230	-	79.3	0	100	NR	206.9	NR	0
Giuliano 2013												
- M	19	74	6	3977	76.0	98.2	20.4	99.8	NR	22.8	NR	4.7
- M + ABUS	42	19	1	3365	97.7	99.7	80.8	99.9	15.2	15.2	15.2	12.3

*0 by design. These studies do not have follow-up (or have limited follow-up) and so are unable to detect the interval cancers over the next year that represent the false negatives. Thus sensitivity and the NPV will always be 100% and the specificity will be overestimated.

Table A3: Characteristics of the screen-detected cancers.

			Mammogram					Supplemental		
Study	N	Size, mm	≤ 1 cm, %	Lymph node negative, %	Stage 0 or 1, %	N	Size, mm	≤ 1 cm, %	Lymph node negative, %	Stage 0 or 1, %
MRI										
Kuhl 2014	-	-	-	-	-	11	8.4	NR	63.6	100
HHUS										
Maestro 1998	-	-	-	-	-	2	15	0	NR	NR
Buchberger 2000	142	11.2	NR	NR	NR	32	9.1	~75	NR	NR
Kaplan 2001	NR	-	-	-	-	6	9	66.7	100	66.7
Kolb 2002	94	-	-	-	-	31	NR	NR	NR	NR
Crystal 2003	NR	13.5	NR	NR	NR	7	9.6	57.1	85.7	NR
Leconte 2003	14	NR	NR	NR	NR	11	NR	NR	NR	NR
Brancato 2007	5.8%	NR	NR	NR	NR	2	NR	50	NR	NR
De Felice 2007	8	NR	NR	NR	NR	12	10			
Corsetti 2008	166	NR	36	68	NR	37	NR	65	86	NR
Corsetti 2011	20	NR	56	94	94	32	NR	84	90	90
Korpraphong 2014	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
Hooley 2012	NR	-	-	-	-	3	7	100	NR	NR
Leong 2012	NR	-	-	-	-	2	10	50	100	100
Weigert 2012	NR	-	-	-	-	28	19	24	9	NR
Chae 2013	6	17	NR	83	NR	24	11	NR	95	NR

			Mammogram					Supplemental		
Study	N	Size, mm	≤ 1 cm, %	Lymph node negative, %	Stage 0 or 1, %	N	Size, mm	≤ 1 cm, %	Lymph node negative, %	Stage 0 or 1, %
Girardi 2013	NR	-	-	-	-	22	<15	-	-	-
Parris 2013	NR	-	-	-	-	10	9.7	NR	77	NR
Venturini 2013	10	-	10	50	66.7	2	-	50	50	50
Berg 2012	59	NR	NR	NR	NR	32	10	NR	96	NR
ABUS										
Arleo 2014	-	-	-	-	-	-	-	-	-	-
Kelly 2010	23	NR	30	NR	83	23	NR	61	NR	78
Stoblen 2011	2	NR	100	100	100	0	-	-	-	-
Brem 2014	17	5.2	NR	100	35.3	30	12.9	NR	92.6	50.9
Giuliano 2013	19	22.3	NR	95	-	42	14.3	NR	98	83
DBT										
Ciatto 2013	39	13.7	NR	72	95	20	13.5	NR	80	95
Destounis 2014	2	NR	NR	NR	NR	6	NR	NR	NR	NR
Friedewald 2014	1027	NR	NR	NR	68	950	NR	NR	NR	74
Greenberg 2014	203	NR	NR	NR	19.1	144	NR	NR	NR	24.5
Haas 2013	37	NR	NR	NR	NR	35	NR	NR	NR	NR
Lourenco 2014	NR	NR	NR	NR	NR	NR	NR	NR	NR	NR
McCarthy 2014	49	NR	NR	NR	NR	85	NR	NR	NR	NR
Rose 2013	56	16	NR	93	86	51	16	NR	88	76
Skaane 2013	77	13.2	49.1	83.0	NR	30	12.8	41.3	85.2	NR

Appendix C

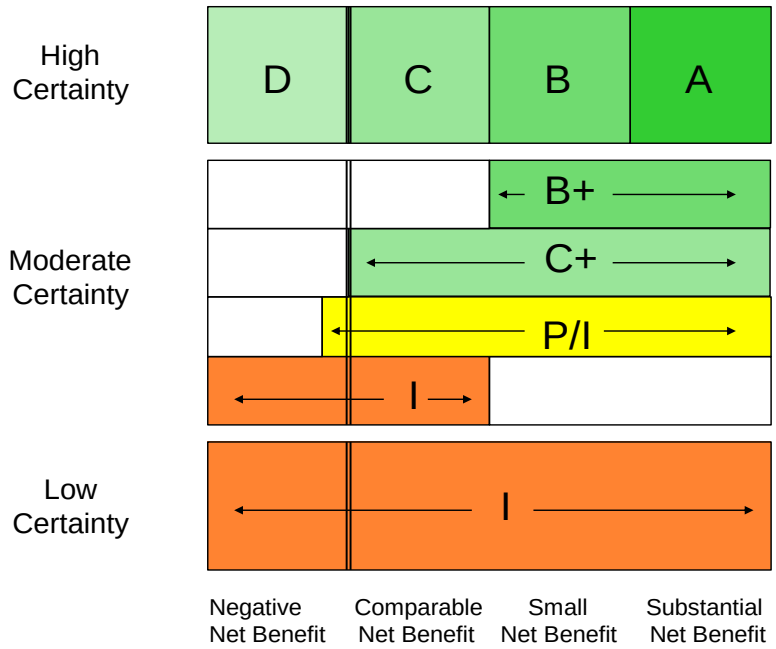
ICER Evidence Matrix

Formulary decisions require a rigorous evaluation of available evidence, a process that entails judgments regarding the quality of individual clinical studies and, ultimately, an assessment of the entire body of evidence regarding a therapeutic agent. To support this latter step, the Institute for Clinical and Economic Review (ICER) has developed the ICER Evidence Rating Matrix™. This user's guide to the ICER Matrix was developed with funding provided by the Comparative Effectiveness Research Collaborative Initiative (CER-CI), a joint initiative of the Academy of Managed Care Pharmacy, the International Society of Pharmacoeconomics and Outcomes Research, and the National Pharmaceutical Council (http://www.npcnow.org/issue/cer-collaborative-initiative). The ICER Matrix presents a framework for evaluating the comparative benefits and risks of therapies in a consistent, transparent system leading to an evidence rating that can guide coverage and formulary placement decisions. The purpose of this user's guide is to help members of Pharmacy and Therapeutics Committees and other decision-makers understand the approach embodied in the matrix, and to help them apply it in a reliable, consistent fashion.

The updated ICER Evidence Rating Matrix is shown below, with a key to the single letter ratings on the following page. Fundamentally, the evidence rating reflects a joint judgment of two critical components:

a) The **magnitude** of the difference between a therapeutic agent and its comparator in "net health benefit" – the balance between clinical benefits and risks and/or adverse effects (horizontal axis); AND

b) The level of **certainty** that you have in your best point estimate of net health benefit (vertical axis).



The letter ratings are listed below, according to the level of certainty in the best estimate of net health benefit.

High Certainty

A = Superior
B = Incremental
C = Comparable
D = Inferior

Moderate Certainty

B+=Incremental or Better
C+=Comparable or Better
P/I = Promising but Inconclusive
I = Insufficient

Low Certainty

I = Insufficient

Steps in Applying the ICER Evidence Rating Matrix

1. **Establish the specific focus of the comparison to be made and the scope of evidence you will be considering.** This process is sometimes referred to as determining the "PICO" – the **P**opulation, **I**ntervention, **C**omparator(s), and **O**utcomes of interest. Depending on the comparison, it is often helpful to also define the specific **T**ime Horizon and **S**etting that will be considered relevant.

2. **Estimate the magnitude of the comparative net health benefit.** Working from the scope of evidence established, it is important to quantify findings from the body of evidence on specific clinical benefits, risks, and other potentially important outcomes, such as adherence, so you can compare these side-by-side for the therapeutic agent and comparator. Some organizations compare each outcome, risk, etc. separately without using a quantitative measure to try to sum the overall comparative balance of benefits and risks between the therapeutic agent and the comparator. For these organizations the estimate of comparative net health benefit must be made qualitatively. Other organizations summarize the balance of benefits and risks using formal mathematical approaches such as health utility analysis, which generates a quantitative summary measure known as the quality-adjusted life year (QALY). What is most important, however, is full and transparent documentation of your rationale for assigning the magnitude of comparative net health benefit into one of four possible categories:

    - ***Negative:*** the drug produces a net health benefit inferior to that of the comparator
    - ***Comparable:*** the drug produces a net health benefit comparable to that of the comparator
    - ***Small:*** the drug produces a small positive net health benefit relative to the comparator
    - ***Substantial:*** the drug produces a substantial (moderate-large) positive net health benefit relative to the comparator

3. **Assign a level of certainty to the estimate of comparative net health benefit.** Given the strength of the evidence on comparative benefits and risks, a "conceptual confidence interval" around the original estimate of comparative net health benefit can be made, leading you to an assignment of the overall level of certainty in that estimate. Rather than assigning certainty by using a fixed equation weighting different attributes of the body of evidence, we recommend formal documentation of the consideration of 5 major domains related to strength of evidence: (1) Level of Bias—how much risk of bias is there in the study designs that comprise the entire evidence base? (2) Applicability—how generalizable are the results to real-world populations and conditions? (3) Consistency—do the studies produce similar treatment effects, or do they conflict in some ways? (4) Directness—are direct or indirect comparisons of therapies available, and/or are direct patient outcomes measured or only surrogate outcomes, and if surrogate outcomes only, how validated are these measures? (5) Precision—does the overall database include enough robust data to provide precise estimates of benefits and harms, or are estimates/confidence intervals quite broad?

   If you believe that your "conceptual confidence interval" around the point estimate of comparative net health benefit is limited to the boundaries of one of the four categories of comparative net health benefit above, your level of certainty is "high". "Moderate" certainty reflects conceptual confidence interval s extending across two or three categories, and may include drugs for which your conceptual confidence interval includes a small likelihood of a negative comparative net health benefit. When the evidence cannot provide enough certainty to limit your conceptual confidence interval within two to three categories of comparative net health benefit, then you have "low" certainty.

4. **Assign a joint rating in the Evidence Rating Matrix.** The final step is the assignment of the joint rating of magnitude of comparative net health benefit and level of certainty. As shown again in the figure on the following page, when your certainty is **"high,"** the estimate of net benefit is relatively assured, and so there are distinct labels available: an **A** rating indicates a high certainty of a substantial comparative net benefit. As the magnitude of comparative net health benefit decreases, the rating moves accordingly, to **B** (incremental), **C** (comparable), and finally **D**, indicating an inferior or negative comparative net health benefit for the therapeutic agent relative to the comparator.

   When the level of certainty in the point estimate is only **"moderate,"** the summary ratings differ based on the location of the point estimate and the ends of the boundaries of the conceptual confidence interval for comparative net health benefit. The ratings associated with moderate certainty include **B+** (incremental or better), which indicates a point estimate of small *or* substantial net health benefit and a conceptual confidence interval whose lower end does not extend into the comparable range. The rating **C+** (comparable or better) reflects a point estimate of either comparable, small, *or* substantial net health benefit and a lower bound of the conceptual confidence interval that does not extend into the inferior range. These ratings may be particularly useful for new drugs that have been tested using noninferiority trial designs, or those involving modifications to an existing agent to provide adherence or safety advantages.

   Another summary rating reflecting moderate certainty is **P/I** (promising but inconclusive). This rating is used to describe an agent with evidence suggesting that it provides a comparable, small, or substantial net benefit over the comparator. However, in contrast to ratings **B+** and **C+, P/I** is the rating given when the conceptual confidence interval includes a small likelihood that the comparative net health benefit might actually be negative. In our experience the **P/I** rating is a common rating when assessing the evidence on novel agents that have received regulatory approval

with evidence of some benefit over placebo or the standard of care, but without robust evidence regarding safety profiles when used in community practice.

The final rating category is **I** (insufficient). This is used in two situations: (a) when there is moderate certainty that the best point estimate of a drug's comparative net health benefit is comparable, but there is judged to be a moderate-high likelihood that further evidence could reveal that the comparative net health benefit is actually negative; and (b) *any* situation in which the level of certainty in the evidence is **"low,"** indicating that limitations in the body of evidence are so serious that no firm point estimate can be given and/or the conceptual confidence interval for comparative net health benefit extends across all 4 categories. This rating would be a common outcome for assessments of the comparative effectiveness of two active drugs, when there are rarely good head-to-head data available; this rating might also commonly reflect the evidence available to judge the comparative effectiveness of a drug being used for an off-label indication.