HEALTH EVIDENCE REVIEW COMMISSION (HERC)

COVERAGE GUIDANCE: DIGITAL BREAST TOMOSYNTHESIS (3D MAMMOGRAPHY) FOR BREAST CANCER SCREENING IN AVERAGE RISK WOMEN

Approved 3/9/2017

HERC Coverage Guidance

Digital breast tomosynthesis for breast cancer screening in average risk women is not recommended for coverage (weak recommendation).

Note: Definitions for strength of recommendation are provided in Appendix A GRADE Informed Framework Element Description.

RATIONALE FOR DEVELOPMENT OF COVERAGE GUIDANCES AND MULTISECTOR INTERVENTION REPORTS

Coverage guidances are developed to inform coverage recommendations for public and private health plans in Oregon as they seek to improve patient experience of care, population health, and the cost-effectiveness of health care. In the era of the Affordable Care Act and health system transformation, reaching these goals may require a focus on population-based health interventions from a variety of sectors as well as individually focused clinical care. Multisector intervention reports will be developed to address these population-based health interventions or other types of interventions that happen outside of the typical clinical setting.

The HERC selects topics for its reports to guide public and private payers based on the following principles:

- Represents a significant burden of disease or health problem
- Represents important uncertainty with regard to effectiveness or harms
- Represents important variation or controversy in implementation or practice
- Represents high costs or significant economic impact
- Topic is of high public interest

Our reports are based on a review of the relevant research applicable to the intervention(s) in question. For coverage guidances, which focus on clinical interventions and modes of care, evidence is evaluated using an adaptation of the GRADE methodology. For more information on coverage guidance methodology, see Appendix A.

Multisector interventions can be effective ways to prevent, treat, or manage disease at a population level. For some conditions, the HERC has reviewed evidence and identified effective interventions, but has not made coverage recommendations, as many of these policies are implemented in settings beyond traditional health care delivery systems.
GRADE-INFORMED FRAMEWORK

The HERC develops recommendations by using the concepts of the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) system. GRADE is a transparent and structured process for developing and presenting evidence and for carrying out the steps involved in developing recommendations. There are several elements that determine the strength of a recommendation, as listed in the table below. The HERC reviews the evidence and makes an assessment of each element, which in turn is used to develop the recommendations presented in the coverage guidance box. Estimates of effect are derived from the evidence presented in this document. The level of confidence in the estimate is determined by the Commission based on the assessment of two independent reviewers from the Center for Evidence-based Policy. Unless otherwise noted, estimated resource allocation, values and preferences, and other considerations are assessments of the Commission.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Estimate of Effect for Outcome/Confidence in Estimate</th>
<th>Resource allocation</th>
<th>Values and Preferences</th>
<th>Other considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>All-cause mortality (Critical outcome)</td>
<td>No data</td>
<td>Based on Medicare fee-for-service fee schedules, DBT increases the cost of mammography by 41%. While the cost of DBT in addition to DM is relatively modest at an individual level, this would add significant costs at the population level due to the large number of people electing breast cancer screening. If DBT+DM leads to lower rates of recall and/or detects breast cancer, women would strongly value having a test that is precise in that it detects cancer that will impact future morbidity and mortality, but would also decrease their risk of unnecessary worry and procedures. If a test is much more likely to pick up a cancer, they would strongly favor it if they know it will affect their long-term outcomes.</td>
<td>Women would strongly value having a test that is precise in that it detects cancer that will impact future morbidity and mortality, but would also decrease their risk of unnecessary worry and procedures. If a test is much more likely to pick up a cancer, they would strongly favor it if they know it will affect their long-term outcomes.</td>
<td></td>
</tr>
<tr>
<td>Breast cancer morbidity (Critical outcome)</td>
<td>No data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test performance characteristics (Important outcome)</td>
<td>Sensitivity for Breast Cancer: Italian study DBT+DM: 0.85 (95% CI 0.74 to 0.92) DM: 0.54 (95% CI 0.42 to 0.65) U.S. study DBT+DM: 0.909 DM: 0.906 OR for improved sensitivity 1.03 (95% CI 0.57 to 1.89)</td>
<td>Based on Medicare fee-for-service fee schedules, DBT increases the cost of mammography by 41%. While the cost of DBT in addition to DM is relatively modest at an individual level, this would add significant costs at the population level due to the large number of people electing breast cancer screening. If DBT+DM leads to lower rates of recall and/or detects breast cancer, women would strongly value having a test that is precise in that it detects cancer that will impact future morbidity and mortality, but would also decrease their risk of unnecessary worry and procedures. If a test is much more likely to pick up a cancer, they would strongly favor it if they know it will affect their long-term outcomes. There would be significant variability in breast cancer morbidity.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Coverage question: Should digital breast tomosynthesis (DBT) be recommended for coverage for breast cancer screening?
| Coverage question: Should digital breast tomosynthesis (DBT) be recommended for coverage for breast cancer screening? |
|---|---|---|---|---|
| **Outcomes** | **Estimate of Effect for Outcome/ Confidence in Estimate** | **Resource allocation** | **Values and Preferences** | **Other considerations** |
| Specificity for Breast Cancer: |  |  |  |  |
| Italian study | DBT+DM: 0.97 (95% CI 0.96 to 0.98)  
DM: 0.96 (95% CI 0.95 to 0.97) |  |  |  |
| U.S. study | DBT+DM: 0.913  
DM: 0.897  
OR for improved specificity 1.22 (95% CI 1.16 to 1.28)  
●○○○ (Very low confidence) |  |  |  |

Cancer Detection Rate:  
DBT+DM: 4.6 to 8.9 per 1,000  
DM: 3.5 to 6.3 per 1,000  
Incremental cancer detection rate with DBT in U.S.-based studies:  
-0.8 to 1.9 per 1,000 (11 studies, 5 with statistically significant increases in incremental cancer detection rate)  
●●○○ (Low confidence)  

For women with dense breasts, incremental cancer detection rate with DBT:  
1.4 per 1,000 (95% CI 0.9 to 2.0) in retrospective studies and 3.9 per 1,000 (95% CI 2.7 to 5.1) in prospective studies  
●●○○ (Low confidence)  

cancer at an earlier stage, leading to improved outcomes, these costs could be offset.  
how women would value an increased risk of a false-positive test and the subsequent need for biopsy or recall compared to a possible missed cancer diagnosis, but we assume that many women would have a strong preference to avoid a missed cancer diagnosis.
**Coverage question:** Should digital breast tomosynthesis (DBT) be recommended for coverage for breast cancer screening?

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Estimate of Effect for Outcome/ Confidence in Estimate</th>
<th>Resource allocation</th>
<th>Values and Preferences</th>
<th>Other considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer stage at diagnosis (Important outcome)</td>
<td>No data</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Recall rate/False-positive rate (Important outcome)** | **Recall rate:**  
DBT+DM: 36 to 136 per 1,000  
DM: 42 to 162 per 1,000  
Change in recall rate with DBT in U.S.-based studies:  
-1.2% to -7.2% (11 studies, all with statistically significant results)  
Biopsy rate:  
DBT+DM: 12 to 27 per 1,000  
DM: 14 to 22 per 1,000  
*PPV Recall:*  
DBT+DM: 4.6% to 29.1%  
DM: 3.0% to 28.5%  
*PPV Biopsy:*  
DBT+DM: 22.7% to 50%  
DM: 16.7% to 30.2%  
●●○○ (Low confidence) |                     |                        |                      |                      |
Coverage question: Should digital breast tomosynthesis (DBT) be recommended for coverage for breast cancer screening?

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Estimate of Effect for Outcome/ Confidence in Estimate</th>
<th>Resource allocation</th>
<th>Values and Preferences</th>
<th>Other considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>For women with dense breasts:</strong>&lt;br&gt;Recall reduction with DBT of 23.3 per 1,000 (95% CI 16.8 to 29.9)&lt;br&gt;●●○○ (Low confidence)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Balance of benefits and harms:** There may or may not be a benefit in terms of cancer detection rate. There does appear to be an improvement in the recall rate. There are no clear harms, unless the possible increased cancer detection rate is mostly due to the detection of indolent and non-invasive lesions.

**Rationale:**
Based on observational studies performed in the US, it is likely that DBT decreases recall rates as compared with DM alone. We have low confidence that DBT improves cancer detection rates. We are also not confident that any improvement in cancer detection rates with DBT, if clearly demonstrated, would result in cancers being detected at earlier stages, leading to earlier intervention that improve clinical outcomes. Adding DBT to standard DM adds cost, and we are not confident that DBT is cost-effective, based on current analysis. Additionally, randomized controlled trials are currently underway that should help with greater understanding of the risks and benefits of DBT+DM, including the critical issue of whether DBT improves clinical outcomes. The recommendation against coverage is a weak recommendation because further evidence could change the recommendation.

**Recommendation:** Digital breast tomosynthesis for breast cancer screening is not recommended for coverage (weak recommendation).

Note: GRADE framework elements are described in Appendix A. The Quality of Evidence rating was assigned based on the GRADE Evidence Profile found in Appendix B.
EVIDENCE OVERVIEW

Clinical Background

Approximately 1 in 8 (12%) women in the United States develop invasive breast cancer during their lifetime, making breast cancer the second most common cancer (following skin cancer) in American women (American Cancer Society [ACS], 2016c). In 2013, there were 230,815 breast cancer diagnoses and 40,860 breast cancer deaths in women in the United States (Centers for Disease Control and Prevention [CDC], 2016a). In men, breast cancer is relatively rare, accounting for an additional 2,109 breast cancer diagnoses and 464 breast cancer deaths in 2013.

Trends in breast cancer incidence and mortality reveal health disparities across race and ethnicity. The rate of breast cancer diagnoses has remained stable in white women over the last decade, while increasing slightly in African American women (ACS, 2016c). The breast cancer mortality rate overall has steadily declined since 1989, but this trend disproportionately represents a larger decrease in breast cancer deaths among white women compared to other races and ethnicities (CDC, 2012). African American women are 40% more likely to die of breast cancer than white women, which reflects the need for more timely follow-up and improved access to high-quality treatment following a positive screening in this population.

Indications

The declining breast cancer mortality rate in the United States is partially attributed to greater screening efforts and thus earlier detection, in addition to fewer women using hormone therapy after menopause and improved quality of treatment (ACS, 2016c). Screening technology, such as mammography, can identify cancer at an earlier stage, before an individual experiences symptoms (ACS, 2016b). When detected early, abnormal tissue or cancer is easier to treat and patients have better outcomes. Women diagnosed with breast cancer in earlier stages have higher relative five-year survival rates (ACS, 2016a). The five-year survival rate for women with Stage 0 or Stage I breast cancer in the United States is almost 100%, compared to 22% for women with Stage IV breast cancer.

A systematic review and meta-analysis of randomized controlled screening trials for breast cancer completed in 2016 to inform the U.S. Preventive Services Task Force (USPSTF) concluded that screening mammography reduces breast cancer mortality, but not all-cause mortality (Nelson et al., 2016b). The absolute reduction in breast cancer mortality afforded by screening mammography varies by age group; for women 39-49 years old screening prevents 3 breast cancer deaths per 10,000 women over 10 years (a finding that was not statistically significant), while in women aged 60-69 years screening prevent 21 deaths per 10,000 women screened over 10 years. The review also concluded that the rate of false-positive recall from screening mammography is high: the cumulative rate of false-positive recalls over 10 years was 61% among women undergoing annual screening and 42% for women receiving biennial screening. On the basis of this review, the USPSTF offered a B recommendation to biennial mammography for average risk women between the ages of 50-74 years and a C recommendation for screening mammography for women 40-49 years old. The USPSTF issued an I recommendation to digital breast tomosynthesis.
The benefits of screening generally increase with age: the greatest benefit is for women aged 50 to 74 (U.S. Preventative Services Task Force, 2016). However, screening recommendations vary by individual case and risk level. Multiple factors contribute to individual risk aside from being female, including age, genetic mutations, denser breasts, family history of breast cancer, physical inactivity, and alcohol consumption (CDC, 2016b).

Advocates of DBT+DM generally recommend it to reduce false-positives and increase cancer detection rate.

**Technology Description**

Digital breast tomosynthesis (DBT), sometimes referred to as three-dimensional (3-D) mammography, is a breast cancer screening technique that was developed to improve detection and characterization of abnormal tissue in the breasts, especially in women with denser breasts (Helvie, 2011). DBT provides multiple X-ray images of thin breast sections, which can potentially reveal cancers concealed by normal tissue. An X-ray tube moves in an arc around a patient’s compressed breast, which allows exposures from different angles to create a series of images. The image dataset is reconstructed into multiple images using mathematical algorithms and then reviewed by a radiologist. This process is distinct from standard (two-dimensional or 2-D) digital mammography (DM), in which only one image of overlapping tissue is produced. The first DBT system was approved by the U.S. Federal Drug Administration (FDA) on February 11, 2011 (U.S. FDA, 2015).

**Key Questions and Outcomes**

The following key questions (KQ) guided the evidence search and review described below. For additional details about the review scope and methods, please see Appendix C.

1. What is the comparative effectiveness of DBT as a primary screening modality in women referred for breast cancer screening?
2. Does the comparative effectiveness of DBT vary by the following characteristics:
   a. Age
   b. Race or ethnicity
   c. Breast density
3. In a screening population, how do the test characteristics of 3-D/2-D mammography compare to those of standard 2-D mammography?
4. What are the harms of 3-D/2-D mammography compared to standard 2-D mammography alone?
5. If DBT is used as a primary screening modality, what is the optimal screening interval, and does that interval vary according to the characteristics listed in Key Question 2?

Critical outcomes selected for inclusion in the GRADE table are all-cause mortality and breast cancer morbidity. Important outcomes selected for inclusion in the GRADE table are test performance characteristics, cancer stage at diagnosis, and recall rate/false-positive test results.
Evidence Review

No randomized controlled trials of DBT have been published, although several are currently underway. Staff identified four recent, high-quality systematic reviews of observational trials of DBT combined with DM compared to DM alone. The included systematic reviews are summarized in Table 1.
Table 1. Summary of Included Systematic Reviews

<table>
<thead>
<tr>
<th>Citation, Study Details</th>
<th>Center QA</th>
<th># of Studies (k), Population (n)</th>
<th>Study Summary and Findings</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Melnikow (2016a)</td>
<td>Good</td>
<td>k = 1 total n = 7,292</td>
<td>Comparators</td>
<td>Included studies</td>
</tr>
<tr>
<td>Search Dates</td>
<td></td>
<td></td>
<td>DBT + DM vs. DM</td>
<td>Ciatto et al., (2013), Houssami et al. (2014) (2 reports from the same study - STORM)</td>
</tr>
<tr>
<td>January 2000 to October 2015</td>
<td></td>
<td></td>
<td>Outcomes</td>
<td>Summarized in evidence tables¹</td>
</tr>
<tr>
<td>Included Study Designs</td>
<td></td>
<td></td>
<td>Sensitivity for Breast Cancer: 0.85 (95% CI, 0.74 to 0.92) vs. 0.54 (95% CI, 0.42 to 0.65)</td>
<td>Destounis et al., (2014), Friedewald et al., (2014), Greenberg et al., (2014), Haas et al., (2013), Lang et al., (2015), McCarthy et al., (2014), Rose et al., (2013), Skaane et al., (2013a)</td>
</tr>
<tr>
<td>Prospective cohort</td>
<td></td>
<td></td>
<td>Specificity for Breast Cancer: 0.97 (95% CI, 0.96 to 0.98) vs. 0.96 (95% CI, 0.95 to 0.97)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citation, Study Details</th>
<th>Center QA</th>
<th># of Studies (k), Population (n)</th>
<th>Study Summary and Findings</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nelson (2016a)</td>
<td>Good</td>
<td>k = 5 total n = 517,011</td>
<td>Comparators</td>
<td>Included studies</td>
</tr>
<tr>
<td>Through December 2014</td>
<td></td>
<td></td>
<td>Outcomes</td>
<td>Evidence limited by lack of RCTs, comparability of results not reported, and outcomes not reported uniformly</td>
</tr>
<tr>
<td>Included Study Designs</td>
<td></td>
<td></td>
<td>Recall Rate: Significantly lower for DBT+ DM vs. DM across studies</td>
<td></td>
</tr>
<tr>
<td>SRs, RCTs, observational studies</td>
<td></td>
<td></td>
<td>One U.S. study reported 16 fewer recalls per 1,000 screens (p&lt;0.001) (Friedewald et al., 2014)</td>
<td></td>
</tr>
</tbody>
</table>

¹ These studies did not meet the inclusion criterion of describing test performance characteristics, but were included in evidence tables to illustrate more proximal outcomes.

Digital Breast Tomosynthesis (3D Mammography) for Breast Cancer Screening in Average Risk Women

Approved 3/9/2017
<table>
<thead>
<tr>
<th>Citation, Study Details</th>
<th>Center QA</th>
<th># of Studies (k), Population (n)</th>
<th>Study Summary and Findings</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>WA HTA (2014)</td>
<td>Good</td>
<td>k = 9 total n = 313,298</td>
<td><strong>Biopsy Rate</strong>: Increase of 1.3 biopsies per 1,000 screens for DBT+ DM compared to DM (p&lt;0.001) (Friedewald et al., 2014)</td>
<td>Included studies Ciatto et al., (2013), Destounis et al., (2014), Friedewald et al., (2014), Greenberg et al., (2014), Haas et al., (2013), Lourenco et al., (2014), McCarthy et al., (2014), Rose et al., (2013), Skaane et al., (2013a), Skaane et al., (2013b) All included articles were rated by the review authors as poor quality due to insufficient follow-up in all but one study, and a 20% dropout rate in the study with 12-month follow-up (Destounis et al., 2014) Some of the studies had possible selection bias Authors reported a moderate to high degree of uncertainty in recall rate, biopsy rate, and CDR There is a low to moderate degree of uncertainty for the PPV of biopsy</td>
</tr>
<tr>
<td>Search Dates</td>
<td></td>
<td></td>
<td><strong>Comparators</strong></td>
<td></td>
</tr>
<tr>
<td>January 1990 to November 2014</td>
<td></td>
<td></td>
<td><strong>DBT+ DM vs. DM</strong></td>
<td></td>
</tr>
<tr>
<td>Included Study Designs</td>
<td></td>
<td></td>
<td><strong>Outcomes</strong></td>
<td></td>
</tr>
<tr>
<td>Observational studies</td>
<td></td>
<td></td>
<td><strong>Cancer Detection Rate (CDR)</strong>: 4 to 6 / 1,000 vs. 3 to 5/1,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Recall Rate</strong>: 80 to 140/1,000 vs. 100 to 160/1,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>Biopsy Rate</strong>: 12 to 27/1,000 vs. 14 to 22/1,000</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><strong>PPV Biopsy</strong>: 25 to 30% vs. 20 to 25%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>*Meta-analysis not performed for outcomes, significance not reported</td>
<td></td>
</tr>
<tr>
<td>Citation, Study Details</td>
<td>Center QA</td>
<td># of Studies (k), Population (n)</td>
<td>Study Summary and Findings</td>
<td>Comments</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------</td>
<td>----------------------------------</td>
<td>-----------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Melnikow (2016b)</td>
<td>Good</td>
<td>k = 4 total n = 30,195</td>
<td>Comparators DBT+DM vs. DM for screening women with dense breasts</td>
<td>Included studies Ciatto et al., (2013), Haas et al., (2013), McCarthy et al., (2014), Rose et al., (2013) The 3 U.S. studies were single-site retrospective designs, and one study included women at above-average risk</td>
</tr>
<tr>
<td>Search Dates March 2000 to July 2015</td>
<td>Observational studies</td>
<td>SR's quality assessment of individual studies: Fair</td>
<td>Outcomes Cancer Detection Rate (CDR): 5.4 to 6.9 / 1,000 vs. 4.0 to 5.2 / 1,000 Recall Rate: 7% to 11% vs. 9% to 17%</td>
<td></td>
</tr>
<tr>
<td>Included Study Designs</td>
<td></td>
<td>*Reported ranges, meta-analysis not performed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search Dates January 200 to July 2016</td>
<td>Retrospective and prospective cohorts</td>
<td></td>
<td>Outcomes Incremental Cancer Detection Rate (CDR): 1.4 to 3.9 more cancers detected per 1,000 with DBT+DM</td>
<td></td>
</tr>
<tr>
<td>Included Study Designs</td>
<td></td>
<td>Recall Rate Reduction: 23.3 fewer recalls per 1,000 with DBT+DM</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**EVIDENCE SUMMARY**

**Key Questions 1 and 3**

*What is the comparative effectiveness of digital breast DBT as a primary screening modality in women referred for breast cancer screening?*

*In a screening population, how do the test characteristics of 3D/2D mammography compare to those of standard 2D mammography?*

*Melnikow et al., 2016a*

A good-quality systematic review of DBT for breast cancer screening conducted for the Agency for Health Research and Quality (AHRQ) was published in January 2016 (Melnikow et al., 2016a). The systematic review included articles published between January 2000 and October 2015 that reported on the test performance of DBT in a screening population (asymptomatic women 40 years of age or older) compared to a comprehensive reference standard that was applied to all test results. The authors of the systematic review required that studies report one year of clinical follow-up after the initial imaging in order to ascertain interval breast cancers that were not detected during screening.

Only a single study met the inclusion criteria. That study, known as the Screening with Tomosynthesis OR standard Mammography (STORM) trial (Houssami et al., 2014), included a prospective cohort of more than 7,000 women aged 48 years or older from northern Italy. These women had both DM and DBT performed at the time of screening. Sequential reading was performed by eight radiologists who read the DM first, then interpreted the combined DM and DBT images. Median follow-up after screening was approximately 20 months. Among this cohort, 63 women were diagnosed with 65 breast cancers during the follow-up period. The authors of the AHRQ review reported the test characteristics from the single-reader analysis because they considered it the most consistent with the practice in the United States. DBT combined with DM (DBT+DM) was more sensitive than DM alone (85% vs. 54%). The two tests had similar specificity (97% for DBT+DM vs. 96% for DM). Overall cancer detection rates were 7.4 per 1,000 for DBT+DM compared to 4.8 per 1,000 for DM. Overall recall rates were 3.6% for DBT+DM compared to 4.2% for DM alone.

The authors of the AHRQ review summarized the results of eight additional screening cohort studies that did not report on test performance, but did report on cancer detection rates, recall rates, and biopsy rates. The authors did not methodologically assess the additional studies, and the results were only summarized in an included evidence table. Overall, the authors concluded that in most studies DBT was associated with increased cancer detection rates, reduced recall rates, and higher positive predictive value for initial recall. The results from the additional trials were mixed with respect to detection of invasive cancers and biopsy rate.
Nelson et al., 2016

A good-quality systematic review of the harm of breast cancer screening, including a comparison of the harms associated with different screening modalities, was published in 2016 to inform the deliberations of the U.S. Preventive Services Task Force (Nelson et al., 2016). The authors included cohort studies performed in asymptomatic populations. The authors assessed the overall quality of the evidence for differential harm by screening modality as poor, noting the absence of randomized trials, the failure to report group characteristics at baseline, and inconsistency in the reporting of outcomes including biopsy rate. Four of the five trials found lower recall rates with DBT+DM compared to DM, and the fifth trial found no significant difference. Only one of the trials reported on biopsy rate and found a statistically significant difference of 1.3 fewer biopsies per 1,000 for DM compared with DBT+DM.

Washington Health Technology Assessment, 2014

A good-quality systematic review of DBT for breast cancer screening was included in a Washington Health Technology Assessment (WA HTA) report released in December 2014 (WA HTA, 2014). The authors included nine studies reported in 10 articles and deemed all of the included studies to be of poor methodological quality. Issues of study heterogeneity prevented a formal meta-analysis, but the authors provided estimations of the cancer detection rate, recall rate, biopsy rate, and positive predictive value of biopsy between the DM and DBT+DM groups. Those results are summarized in Table 2.

Table 2. Summary Comparison of DM and DBT+DM from the WA HTA report

<table>
<thead>
<tr>
<th>Outcome</th>
<th>DM</th>
<th>DBT+DM</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer detection rate (per 1,000)</td>
<td>3–5</td>
<td>4–6</td>
<td>Moderate-high</td>
</tr>
<tr>
<td>Recall rate (per 1,000)</td>
<td>100–160</td>
<td>80–140</td>
<td>Moderate-high</td>
</tr>
<tr>
<td>Biopsy rate (per 1,000)</td>
<td>14–22</td>
<td>12–27</td>
<td>Moderate</td>
</tr>
<tr>
<td>Positive predictive value of biopsy</td>
<td>20–25%</td>
<td>25–30%</td>
<td>Low-moderate</td>
</tr>
</tbody>
</table>
Melnikow et al., 2016b

A good-quality systematic review of supplemental breast cancer screening for women with dense breasts, including DBT, was published in 2016 (Melnikow et al., 2016b). The authors included four fair-quality studies of DBT+DM compared to DM alone in women with dense breasts. The three U.S. studies were single retrospective cohorts comparing outcomes before and after implementation of DBT. None of the included studies reported on test performance characteristics. Three of the studies reported on cancer detection rate (4.0–5.2 per 1,000 for DM compared to 5.4–6.9 per 1,000 for DBT+DM); one of the studies reported that the rate of invasive cancers was the same between the two groups. Among the three U.S. studies included, recall rates were lower for DBT+DM (7%–11%) compared to DM (9%–17%). The authors noted that there is no reference standard by which to measure the accuracy of BI-RADS density determinations and that reclassification of breast density on sequential exams is common.

Results from individual studies included in the systematic reviews or submitted in public comments are summarized in Table 3, and Table 4 summarizes the U.S.-based studies.

Table 3. Results from Individual Studies Included in the Systematic Reviews or Submitted in Public Comment

<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Study Size Location</th>
<th>Study Design and Population Characteristics</th>
<th>CDR per 1,000 women (% invasive)</th>
<th>Recall Rate (%)</th>
<th>PPV Recall</th>
<th>PPV Biopsy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ciatto (2013)</td>
<td>n = 7,292 Italy</td>
<td>Prospective cohort (one arm)</td>
<td>DBT+DM: 8.1 DM: 5.3 (p&lt;0.0001)</td>
<td>DBT+DM: 4.3% DM: 5% (NS)</td>
<td>NR</td>
<td>NR</td>
<td>No long-term follow-up; one abnormal read-flagged recall</td>
</tr>
<tr>
<td>Destounis (2014)</td>
<td>Poor (WA HTA, 2014)</td>
<td>Retrospective cohort (two arm)</td>
<td>DBT+DM: 5.4 (33%) DM: 3.8 (50%)</td>
<td>DBT+DM: 4.2% DM: 11.4%</td>
<td>NR</td>
<td>DBT+ DM: 50.0% DM: 16.7%</td>
<td>One-year follow-up; 80% completion rate</td>
</tr>
<tr>
<td><strong>Author (Year)</strong></td>
<td><strong>Study Size</strong></td>
<td><strong>Location</strong></td>
<td><strong>Source QA</strong></td>
<td><strong>Study Design and Population Characteristics</strong></td>
<td><strong>CDR per 1,000 women (% invasive)</strong></td>
<td><strong>Recall Rate (%)</strong></td>
<td><strong>PPV Recall</strong></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>--------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------</td>
<td>-------------------</td>
<td>---------------</td>
</tr>
</tbody>
</table>
Mean age: 59  
Test: Selenia Dimensions, Hologic  
SecurView, Hologic | (sig. NR) | (p<0.0001) | (sig. NR) | Selection bias likely due to baseline risk factors for breast cancer or abnormal imaging in the DBT group; some participants had a personal history of breast cancer |
| Poor (WA HTA, 2014) | Retrospective cohort (two arm): Pre-post  
13 academic medical centers and breast diagnostic/screening centers  
Mean age: 56.2 for DBT+DM; 57.0 for DM  
Test: Selenia Dimensions, Hologic | DBT+DM: 5.5 (75%)  
DM: 4.3 (67%) (p<0.001) | DBT+DM: 8.9%  
DM: 10.6% (p<0.001) | DBT+DM: 6.1%  
DM: 4.1% (p<0.0001) | DBT+DM: 29.2%  
DM: 24.2% (p<0.001) | Insufficient follow-up  
Pre-post design  
No individual-level data to stratify populations  
The biopsy rate was higher for DBT+DM group: 1.9% vs. 1.8% (p=0.004) |
| Poor (WA HTA, 2014) | Retrospective cohort (two arm) | DBT+DM: 6.3 (74%)  
DM: 4.9 (62%) | DBT+DM: 13.6%  
DM: 16.2% (p<0.0001) | DBT+DM: 4.6%  
DM: 3.0% (p=0.0003) | DBT+DM: 22.7%  
DM: 21.5% (NS) | No follow-up  
Volunteer bias possible |
<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Study Size</th>
<th>Location</th>
<th>Source QA</th>
<th>Study Design and Population Characteristics</th>
<th>CDR per 1,000 women (% invasive)</th>
<th>Recall Rate (%)</th>
<th>PPV Recall</th>
<th>PPV Biopsy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBT+DM exams = 20,943 DM exams = 38,674 Washington, D.C.</td>
<td>Community-based multisite radiology practice Mean age: 59.5 Test: Selenia Dimensions, Hologic</td>
<td>Community-based multisite radiology practice Mean age: 59.5 Test: Selenia Dimensions, Hologic</td>
<td>(p=0.035)</td>
<td>(p=0.035)</td>
<td>May have overlap with Friedewald (2014) DBT+DM group had higher biopsy rate (2.6% vs. 2.1%, p=0.0003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haas (2013) DBT+DM = 6,100 DM = 7,058 Connecticut</td>
<td>Poor (WA HTA, 2014) Retrospective cohort (two arm) Mean age: 56 Test: Selenia Dimensions, Hologic</td>
<td>Retrospective cohort (two arm) Mean age: 56 Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 5.7 (69%) DM: 5.2 (68%) (NS)</td>
<td>DBT+DM: 8.4% DM: 12.0% (p&lt;0.01)</td>
<td>DBT+DM: 6.8% DM: 4.3%² NR</td>
<td>No follow-up Women in DBT group had increased risk factors for breast cancer at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Houssami (2014) n = 7,292</td>
<td>Good (Melnikow, 2016a) Prospective cohort (one arm)</td>
<td>Good (Melnikow, 2016a) Prospective cohort (one arm)</td>
<td>DBT+DM: 7.4 DM: 4.8 (p&lt;0.001)</td>
<td>DBT+DM: 3.6% DM: 4.2% (NS)</td>
<td>DBT+DM: 21% DM: 11%³ NR</td>
<td>Follow-up 13 months or greater Screen positive if one of two readers</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

² Center staff calculated this by dividing cancers detected by the product of the recall rate and the number of exams, significance not reported.
³ Drawn from AHRQ (2016) report; PPV not reported in original study. Significance not recorded.
<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Study Size Location</th>
<th>Source QA</th>
<th>Study Design and Population Characteristics</th>
<th>CDR per 1,000 women (% invasive)</th>
<th>Recall Rate (%)</th>
<th>PPV Recall</th>
<th>PPV Biopsy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td></td>
<td></td>
<td>Population screening program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Median age: 58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>interpreted DM or DBT as abnormal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Test: Selenia Dimensions, Hologic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lourenco</td>
<td>Poor (WA HTA, 2014)</td>
<td>Retrospective cohort (two arm)</td>
<td>DBT: 4.6 DM: 5.4 (NS)</td>
<td>DBT: 6.4% DM: 9.3% (p&lt;0.00001)</td>
<td>DBT: 7.2% DM: 5.8% (NS)</td>
<td>DBT: 23.8% DM: 30.2% (sig. NR)</td>
<td>Insufficient follow-up</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Single breast imaging center</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pre-post design</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean age: 55.3 DBT, 54.6 DM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Biopsy rate 1.7% DBT+DM vs. 1.6% DM (stat dif NR)</td>
</tr>
<tr>
<td>McCarthy</td>
<td>Poor (WA HTA, 2014)</td>
<td>Cohort (two arm)</td>
<td>DBT+DM: 5.5 (71%) DM: 4.6 (69%) (NS)</td>
<td>DBT+DM: 8.8% DM: 10.4% (p&lt;0.001)</td>
<td>DBT+DM: 6.2% DM: 4.4% (p=0.05)</td>
<td>DBT+DM: 25.7% DM: 24.7% (NS)</td>
<td>Insufficient follow-up</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>One academic medical center</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overlap with Friedewald (2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean age: 57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pre-post design</td>
</tr>
<tr>
<td>Author (Year)</td>
<td>Study Size</td>
<td>Location</td>
<td>Source QA</td>
<td>Study Design and Population Characteristics</td>
<td>CDR per 1,000 women (% invasive)</td>
<td>Recall Rate (%)</td>
<td>PPV Recall</td>
<td>PPV Biopsy</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
<td>---------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Rose (2013)</td>
<td>DM exams = 10,728</td>
<td>Pennsylvania</td>
<td>Poor (WA HTA, 2014)</td>
<td>Cohort (two arm) Multisite, community-based Mean age: NR Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 5.4 (80%) DM: 4.0 (70%) (NS)</td>
<td>DBT+DM: 5.5% DM: 8.7% (p&lt;0.001)</td>
<td>DBT+DM: 10.1% DM: 4.7% (p&lt;0.001)</td>
<td>DBT+DM: 39.8% DM: 26.5% (p=0.06)</td>
</tr>
<tr>
<td>Skaane (2013a)</td>
<td>n = 12,621 exams</td>
<td>Norway</td>
<td>Poor (WA HTA, 2014)</td>
<td>Prospective cohort (one arm) Citywide screening program</td>
<td>DBT+DM: 8.0 (80%) DM: 6.1 (73%) (p=0.001)</td>
<td>DBT+DM: 6.1% DM: 6.7% (p&lt;0.001)</td>
<td>DBT+DM: 29.1% DM: 28.5% (NS)</td>
<td>NR</td>
</tr>
<tr>
<td>Author (Year)</td>
<td>Study Size</td>
<td>Location</td>
<td>Source QA</td>
<td>Study Design and Population Characteristics</td>
<td>CDR per 1,000 women (% invasive)</td>
<td>Recall Rate (%)</td>
<td>PPV Recall</td>
<td>PPV Biopsy</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------</td>
<td>----------</td>
<td>-----------</td>
<td>---------------------------------------------</td>
<td>---------------------------------</td>
<td>-----------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>Skaane (2013b)</td>
<td>n = 12, 621 exams</td>
<td>Norway</td>
<td>Poor (WA HTA, 2014)</td>
<td>Prospective cohort (one arm) Citywide screening program Mean age: 59.3 Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 9.4 DM: 7.1 (p&lt;0.001)</td>
<td>DBT+DM: 3.7% DM: 2.9% (p&lt;0.001)</td>
<td>DBT+DM: 24.7% DM: 25.5% (NS)</td>
<td>NR</td>
</tr>
</tbody>
</table>

Mean age: NR Test: Selenia Dimensions, Hologic

CDR and recall rate calculated for each image prior to arbitration

Independent double-reading with arbitration prior to recall
<table>
<thead>
<tr>
<th>Author (Year)</th>
<th>Study Size Location</th>
<th>Source QA</th>
<th>Study Design and Population Characteristics</th>
<th>CDR per 1,000 women (% invasive)</th>
<th>Recall Rate (%)</th>
<th>PPV Recall</th>
<th>PPV Biopsy</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sharpe (2016)</td>
<td>n = 85,852 exams United States</td>
<td>Poor</td>
<td>Retrospective cohort Single center Mean age: 57.6 Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 5.4 DM: 3.5 (p&lt;0.001) Invasive CDR DBT+DM: 2.81 DM: 2.46 (NS)</td>
<td>DBT+DM: 6.1% DM: 7.5% (p&lt;0.001)</td>
<td>NR</td>
<td>NR</td>
<td>Baseline differences between cohorts Less experienced readers were excluded from recall rate analysis</td>
</tr>
<tr>
<td>Rose (2014)</td>
<td>n = 10,878 United States</td>
<td>Poor</td>
<td>Retrospective reading study Single practice Mean age: NR Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 5.4 (81%) DM: 3.5 (76%) (p&lt;0.001)</td>
<td>DBT+DM: 5.4% DM: 8.2% (p&lt;0.001)</td>
<td>NR</td>
<td>NR</td>
<td>Retrospective reading may influence interpretation</td>
</tr>
<tr>
<td>Author (Year)</td>
<td>Study Size Location</td>
<td>Source QA</td>
<td>Study Design and Population Characteristics</td>
<td>CDR per 1,000 women (% invasive)</td>
<td>Recall Rate (%)</td>
<td>PPV Recall</td>
<td>PPV Biopsy</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
<td>-----------</td>
<td>---------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>McDonald (2016)</td>
<td>n = 44,468 United States</td>
<td>Poor</td>
<td>Retrospective cohort study Single center Mean age: 56.8 Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 5.5 to 6.1 DM: 4.6 (NS)</td>
<td>DBT+DM: 8.8 to 9.2% DM: 10.4% (p&lt;0.001)</td>
<td>DBT+DM: 6.2% to 6.7% DM: 4.4% (p=0.02)</td>
<td></td>
<td>Reported as not statistically significantly different (p=0.37)</td>
</tr>
<tr>
<td>Lang (2016)</td>
<td>n = 7,500 Sweden</td>
<td>Fair</td>
<td>Prospective cohort (one arm) Citywide screening program Mean age: 56 Test: Mammostat Inspiration, Siemens</td>
<td>DBT: 8.9 DM: 6.3 (p&lt;0.0001)</td>
<td>DBT: 3.8% DM: 2.6% (p&lt;0.0001)</td>
<td>DBT: 24% DM: 24% (NS)</td>
<td></td>
<td>NR</td>
</tr>
<tr>
<td>Author (Year)</td>
<td>Study Size Location</td>
<td>Source QA</td>
<td>Study Design and Population Characteristics</td>
<td>CDR per 1,000 women (% invasive)</td>
<td>Recall Rate (%)</td>
<td>PPV Recall</td>
<td>PPV Biopsy</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------------</td>
<td>-----------</td>
<td>--------------------------------------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>------------</td>
<td>------------</td>
<td>----------</td>
</tr>
<tr>
<td>Conant (2016)</td>
<td>n = 142,883 United States</td>
<td>Poor</td>
<td>Retrospective cohort Three centers in the Northeast U.S. Mean age: NR Test: NR</td>
<td>DBT+DM: 5.9 DM: 4.4 (p=0.0026) Invasive CDR DBT+DM: 4.4 DM: 3.3 (p=0.0449)</td>
<td>DBT+DM: 8.7% DM: 10.4% (p&lt;0.0001)</td>
<td>DBT+DM: 6.4% DM: 4.1% (p&lt;0.001)</td>
<td>NR</td>
<td>Baseline differences between cohorts</td>
</tr>
<tr>
<td>Bernardi (2016)</td>
<td>n = 9,672 Northern Italy</td>
<td>Fair</td>
<td>Prospective double-reading study Population-based screening program in Italy Mean age: 58 Test: Selenia Dimensions, Hologic</td>
<td>DBT+DM: 8.8 DM: 6.3 (p&lt;0.0001)</td>
<td>False-positive recall rate: DBT+DM 4.45% DM: 3.42% (p&lt;0.0001)</td>
<td>NR</td>
<td>NR</td>
<td>Double reading model is uncommon in the U.S. Inadequate follow-up to detect interval cancers</td>
</tr>
</tbody>
</table>
Table 4. Incremental Cancer Detection Rate and Change in Recall Rate for DBT+DM over DM alone in U.S.-based studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Incremental Cancer Detection Rate</th>
<th>Change in Recall Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haas (2013)</td>
<td>0.5 per 1,000 (NS)</td>
<td>-3.6%</td>
</tr>
<tr>
<td>Rose (2013)</td>
<td>1.4 per 1,000 (NS)</td>
<td>-3.2%</td>
</tr>
<tr>
<td>Destounis (2014)</td>
<td>1.6 per 1,000 (NR)</td>
<td>-7.2%</td>
</tr>
<tr>
<td>Friedewald (2014)</td>
<td>1.2 per 1,000</td>
<td>-1.7%</td>
</tr>
<tr>
<td>Greenberg (2014)</td>
<td>1.4 per 1,000</td>
<td>-2.6%</td>
</tr>
<tr>
<td>Lourenco (2014)</td>
<td>-0.8 per 1,000 (NS)</td>
<td>-2.9%</td>
</tr>
<tr>
<td>McCarthy (2014)</td>
<td>0.9 per 1,000 (NS)</td>
<td>-1.6%</td>
</tr>
<tr>
<td>Rose (2014)</td>
<td>1.9 per 1,000</td>
<td>-2.8%</td>
</tr>
<tr>
<td>Conant (2016)</td>
<td>1.5 per 1,000</td>
<td>-2.3%</td>
</tr>
<tr>
<td>McDonald (2016)</td>
<td>1.5 per 1,000 (NS)</td>
<td>-1.2%</td>
</tr>
<tr>
<td>Sharpe (2016)</td>
<td>1.9 per 1,000</td>
<td>-1.4%</td>
</tr>
</tbody>
</table>

All observed differences were statistically significant except where noted by NS (not significant) or NR (test of statistical significance not reported). For additional details, see Table 3.

**Critical Outcome: All-cause mortality**

None of the identified studies reported on the effects of DBT on all-cause mortality.

**Critical Outcome: Breast cancer morbidity**

None of the identified studies reported on the effects of DBT on breast cancer morbidity.

**Important Outcome: Test performance characteristics**

Only a single study from the included systematic reviews was designed to allow estimation of the test performance characteristics in a screening population (Houssami et al., 2014). In this prospective study of single-reader breast cancer detection that followed women for a median of nearly 20 months to detect interval cancers, the sensitivity of DBT+DM (0.85, 95% CI 0.74 to 0.92) was superior to that of DM (0.55, 95% CI 0.42 to 0.65). Specificity for DBT+DM was 0.97 (95% CI 0.96 to 0.98) compared to 0.96 (95% CI 0.95 to 0.98) for DM. As the AHRQ review authors noted, the observed sensitivity for DM in this study was well below that reported in other studies.

All 11 of the studies included in the systematic reviews reported on cancer detection rate. Five studies found significantly higher cancer detection rates for DBT+DM compared to DM. Four studies found no
significant differences in cancer detection rate, and one study did not report a test of statistical significance for the outcome.

**Important Outcome: Cancer stage at diagnosis**

None of the identified studies reported on the effects of DBT on cancer stage at diagnosis. Seven of the studies included in the systematic reviews reported on the percentage of detected cancers that were deemed invasive. For DBT+DM, 33% to 80% of the detected cancers were invasive compared to 50% to 74% for DM alone. Most of these studies reported similar or slightly higher rates of invasive disease among the cancers detected by DBT+DM compared to DM.

**Important Outcome: Recall rate/false-positive test results**

Among 11 studies included in the systematic reviews, nine found statistically significantly lower recall rates with DBT+DM compared to DM, and two found no difference. The reported recall rates ranged from 3.6% to 13.6% for DBT+DM and 4.2% to 16.2% for DM. The summary estimate of recall rates provided in the WA HTA report was 80–140 per 1,000 for DBT+DM and 100–160 per 1,000 for DM. The WA HTA report also found similar biopsy rates between the two groups (12–27 per 1,000 for DBT+DM vs. 14–22 per 1,000 for DM alone). In the WA HTA report, the estimate of the positive predictive value of biopsies indicated by DBT+DM was higher than that for biopsies indicated by DM alone (25–30% vs. 20–25%).

In the STORM study (Ciatto et al., 2013), the overall false-positive recall rate was 5.5%, with a significantly greater number of false-positive recalls attributable to DM (n=141) compared to DBT+DM (n=73). In the Oslo study (Skaane et al., 2013a), the overall false-positive recall rate for DBT+DM was lower than that of DM (5.3% vs. 6.1%, p<0.01).

Among the studies included in the systematic reviews, four studies found statistically significant increases in the positive predictive value of recall for DBT+DM compared to DM; three studies (two of which were conducted in Europe where overall recall rates are lower and the positive predictive value of recall is higher) found no significant differences in the positive predicative value of recall, and four studies either did not report on that outcome or did not report tests of statistical significance.

**Key Question 2**

*Does the comparative effectiveness of DBT vary by the following characteristics:*

  a. Age  
  b. Race or ethnicity  
  c. Breast density

The STORM study (Ciatto et al., 2013) reported on cancer detection rate by age group and breast density. The incremental cancer detection rate of DBT+DM compared to DM was 1.7 per 1,000 among women under age 60 compared to 4.0 per 1,000 in women aged 60 years or older. The incremental cancer detection rate was similar among women with lower breast density (2.8 per 1,000) and higher breast density (2.5 per 1,000), although the authors cautioned that the small number of women with higher breast density limits the comparison.
In the retrospective study by Haas and colleagues (2013), DBT+DM was associated with statistically significant reductions in recall rates for women in all age groups with the exception of those women 70 years of age or older. The authors also reported statistically significant reductions in recall rates for women with any breast density classification other than predominantly fatty.

The systematic review by Melnikow and colleagues on screening for women with dense breasts (2016b) found that in three studies the cancer detection rate was superior with DBT+DM (5.4–6.9 per 1,000) compared to DM (4.0–5.2 per 1,000), with one study also demonstrating equivalent proportions of invasive cancers in both groups. The reported recall rates were also lower with DBT+DM (range 7% to 11%) compared to DM (9% to 17%).

**Key Question 4**

*What are the harms of 3-D/2-D mammography compared to standard 2-D mammography alone?*

Overdiagnosis occurs when noninvasive or indolent cancers that would not cause morbidity are identified during screening. Estimates of the percentage of breast cancers that represent overdiagnosis range from approximately 10% to 20% based on randomized controlled trials of mammography (Nelson et al., 2016a). None of the included studies of DBT specifically reported on overdiagnosis, although as noted above, the studies that reported on invasive cancers found similar proportions between DM and DBT+DM groups.

False-positive results have been associated with higher levels of breast-cancer worry and distress in three fair- to good-quality systematic reviews, but the effects of false-positive tests on screening reattendance, anxiety, and depression were mixed (Nelson et al., 2016). As noted above, the overall recall rate for DBT+DM is similar to, or slightly lower than, the recall rate for DM alone. Four studies found statistically significant improvements in the positive predictive value of initial recall with DBT+DM while two studies found no significant difference.

Estimates of the incidence of radiation-induced cancer death from mammography vary based on age and screening interval, but range from 2 per 100,000 to 11 per 100,000 (Nelson et al., 2016). DBT and DM require similar doses of radiation. When DBT and DM images are acquired separately, the dose of radiation is effectively doubled (Melnikow et al., 2016a). Technology that became available in 2013 allows reconstruction of two-dimensional images, thus limiting the radiation dose to that required for a single examination.

**Key Question 5**

*If DBT is used as a primary screening modality, what is the optimal screening interval, and does that interval vary according to the characteristics listed in Key Question 2?*

Because none of the studies followed participants beyond one year or through subsequent rounds of screening with DBT, the optimal screening interval with DBT cannot be established from the existing evidence.
Additional Studies

Sharpe et al., 2016
This is a single center retrospective cohort study of 85,852 examinations for asymptomatic women presenting for breast cancer screening between 2011 and 2014. In the overall cohort, there were 5,703 DBT+DM examinations (6.6%) and 80,149 DM examinations. Nearly all of the examinations (89%) were interpreted by 10 radiologists specializing in breast imaging; the remaining 11% were interpreted by lower volume general radiologists and these examinations were excluded from the recall rate analysis. There were baseline differences in patient characteristics between the DBT cohort and DM cohort. In the DBT cohort there were greater proportions of women with extremely dense breasts, family history of breast cancer, personal history of breast cancer, and a history of breast biopsy with benign findings compared to the DM cohort.

The cancer detection rate in the DM cohort was 3.5 per 1,000 (95% CI 3.1 to 3.9) compared to 5.4 per 1,000 (95% CI 3.7 to 7.8), resulting in an incremental cancer detection rate of 1.9 per 1,000 (p<0.0018). However, the invasive cancer detection rate was 2.46 per 1,000 in the DM cohort and 2.81 per 1,000 in the DBT+DM cohort, a difference that was not statistically significant (p=0.61). The detection rate for carcinoma in situ was 1.04 per 1,000 in the DM cohort and 2.63 per 1,000 in the DBT+DM cohort (p<0.0006).

The overall recall rate was statistically significantly lower in the DBT+DM cohort (6.10%) compared to the DM cohort (7.51%) (p<0.0001). Subgroups in which there were statistically significant differences in the recall rate (in favor of DBT) included women with heterogeneously or extremely dense breasts and women in their fifth and seventh decades. However, in a mixed-effects logistic regression model intended to account for other factors associated with differences in recall rates, the authors found that there was no association between the type of study and the risk of recall after the other factors were accounted for (p=0.7459).

Rose et al., 2014
This is a retrospective reading study in which seven radiologists from a single practice retrospectively interpreted the full-field DM images obtained from 10,878 DM+DBT examinations that had been prospectively interpreted by radiologists in a clinical setting. The retrospective readers interpreted the DM images at least 120 days after the examination was performed, and they were blinded to the original clinical interpretations and subsequent clinical course.

Of the 10,878 DBT+DM examinations that were prospectively interpreted, there were 588 recalls (5.4%) and 59 cancers detected (5.4 per 1,000, 81% invasive). In the retrospective reading of DM images alone, there would have been 888 recalls (8.2%) and 38 cancers detected (3.5 per 1,000, 76% invasive). For examinations that were indicated for recall only by retrospective review of the DM images, a third radiologist performed arbitration, and only 3 of these women were ultimately given a late recall; one of the late recalls resulted in a diagnosis of infiltrating lobular carcinoma and one resulted in a diagnosis of ductal carcinoma in situ. Of the 362 instances in which recall was indicated by DBT+DM but not by DM alone, there were 21 cancers detected, of which 19 were invasive. In the overall group, there was one case of breast cancer that was not screen detected over the course of one year of follow-up.
McDonald et al., 2016
This is a single center retrospective cohort study of 44,468 screening examination in 23,958 unique asymptomatic women between 2010 and 2014. In the first year (September 2010 to September 2011), 10,728 screening exams were performed using DM (denoted DM0 cohort). Between October 2011 and October 2014, 33,740 screening exams were performed using DBT (denoted DBT1, DBT2, or DBT3 cohorts based on the year the exam was performed). Information about the clinical course was taken from the medical record and the Pennsylvania State Cancer Registry (to assess the rate of interval cancers). There were no differences in the baseline patient characteristics between the DM0 and DBT3 cohorts.

The cancer detection rate was 4.6 per 1,000 in the DM0 cohort compared with 5.5 per 1,000 in the DBT1, 5.8 per 1,000 in the DBT2, and 6.1 per 1,000 in the DBT3 cohorts; none of the differences between the cancer detection rate in the DM0 cohort and each of the DBT cohorts were statistically significant. Similarly, there was no statistically significant difference in the rate of invasive cancer detection between the DM0 cohort and any of the DBT cohorts.

The recall rate in the DM0 cohort was 104 per 1,000 compared to 88 per 1,000, 90 per 1,000, and 92 per 1,000 in the DBT1, DBT2, and DBT 3 cohorts respectively. The difference in the recall rate was statistically significant between the DM0 cohort and each of the DBT cohorts. The overall biopsy rate was about 2% and did not differ between the cohorts. Notably, the recall rate for the most recent screening for women receiving serial rounds of DBT screening fell with each additional exam (78 per 1,000 in women who received two screenings and 59 per 1,000 in women who received three screenings).

The positive predictive value of recall (the number of cancers diagnosed per patients recalled to undergo biopsy, abbreviated PPV1) was 4.4% in the DM0 cohort, 6.2% in the DBT1 cohort, 6.5% in the DBT2 cohort, and 6.7% in the DBT3 cohort; the differences in PPV1 between the DM0 and DBT2 or DBT3 cohorts were statistically significant. The number of cancers per biopsy recommended (PPV2) and the number of cancers per biopsy performed (PPV3) were not statistically significantly different in comparing any of the cohorts. The interval cancer rate during the DM0 year was 0.7 per 1,000 compared to 0.5 per 1,000 in the DM1 year, a difference that was not statistically significant.

Lang et al., 2016
This is a prospective population-based single-arm screening cohort of asymptomatic Swedish women between the ages of 40 and 74 years. Among all women presenting to the screening program, 7,500 were randomly selected to undergo one-view DBT and two-view DM with independent, blinded double reading of the exams (double reading is a common practice in much of Europe). Six experienced readers interpreted all of the examinations; if either of the studies (DBT or DM) was interpreted as positive by one reader, that study was referred for arbitration by two other readers who made a final decision on recall. All of the follow-up evaluation for recalled patients was done at the same center. The average age of participating women was 56 years, and about 20% were undergoing their first screening examination; 34% had heterogeneously dense breasts and 8% had dense breasts.
Cancer was identified in 67 women in the DBT reading arm and 47 in the DM reading arms; 46 of the cancers were detected by both modalities, meaning that 21 cancers were detected only by DBT and one cancer was detected only by DM. The overall cancer detection rate was 8.9 per 1,000 (95% CI 6.9 to 11.3) for DBT and 6.3 per 1,000 (95% CI 4.6 to 8.3) for DM. The authors state that DBT offered a statistically significant increase in cancer detection of 43% over DM (95% CI 21% to 68%, p<0.0001). In contrast to most studies that involve sequential interpretation of DBT and DM images, in this study all of the cancers found in the DBT reading arm were identified by DBT images alone.

There were more recalls in the DBT reading arm (3.8%, 95% CI 3.3% to 4.2%) compared to the DM reading arm (2.6%, 95% CI 2.3% to 3.0%) The authors stated that DBT resulted in a statistically significant increase in the recall rate over DM of 43% (95% CI 26% to 62%, p<0.0001). The positive predictive value of recall was the same in both groups (24%). Of the 21 cancers detected in the DBT reading arm but not in the DM reading arm, 81% were invasive. There were no statistically significant differences in the cancer type, histologic grade, tumor size, or lymph node status between the cases detected only in the DBT reading arm and those detected in the DM reading arm, although the study was inadequately powered to detect such differences.

**Conant et al., 2016**

This is a retrospective cohort study involving 142,883 DM exams and 55,998 DBT exams performed at three centers in the Northeast U.S. between 2011 and 2014. The results from one of these centers were previously reported in part (Sharpe et al., 2016). At the second center, DBT+DM was offered based on availability and patient preference; at the third center, DBT+DM was used at the request of patients or providers or was targeted to women with dense breasts, baseline exams, or no previous imaging. Among the entire cohort, a smaller subset (25,268 DBT+DM exams and 113,061 DM exams) had at least 12 months of follow-up that allowed calculation of the overall cancer rate (derived from data from statewide cancer registries) and the sensitivity and specificity of the exams.

There were important baseline differences between the women in the DBT+DM cohort and the DM cohort with respect to age distribution (more younger women in the DBT cohort), race/ethnicity (more black women in the DBT cohort), breast density (more women with heterogeneously dense breasts in the DBT cohort), and first-time screening (more first-time screening exams in the DBT cohort). Additionally, nearly all of the women in the DM cohort came from two of the three study centers. The authors applied a priori logistic regression model meant to account for the differences in age, breast density, first-time screening, and research center.

Recall rates and biopsy rates were calculated for the overall cohort. The recall rate was 8.7% in the DBT+DM group compared to 10.4% in the DM group (odds ratio [OR] for recall in the adjusted analysis 0.68, 95% CI 0.65 to 0.71). The biopsy rate was 2% in the DBT+DM group compared to 1.8% in the DM group, but in the adjusted analysis the odds ratio for biopsy was lower in the DBT+DM group (OR 0.85, 95% CI 0.77 to 0.93). The reduction in recall for DBT+DM was greatest in women between age 40 and 49 and in women with dense breasts, but all subgroups showed a reduction in recall.

The cancer outcomes were calculated in the smaller cohort for individuals for whom at least 12 months of follow-up was available. The observed cancer rate was higher in the DBT+DM group (6.5 per 1,000)
than the DM group (4.9 per 1,000). The invasive cancer rate was 4.7 per 1,000 in the DBT+DM group and 3.7 per 1,000 in the DM group. The cancer detection rate was 5.9 per 1,000 in the DBT+DM group compared with 4.4 per 1,000 in the DM group (p=0.0026). The invasive cancer detection rate was 4.2 per 1,000 in the DBT+DM group and 3.3 per 1,000 in the DM group (p=0.0449). The false-negative rate was 0.60 per 1,000 in the DBT+DM group and 0.46 per 1,000 in the DM group (p=0.347). The positive predictive value of recall was 6.4% in the DBT+DM group and 4.1% in the DM group (p<0.0001). There was no difference in the sensitivity of the two exams (90.6% vs. 90.9%, p=1.00), but the specificity was higher for DBT+DM (91.3%) than for DM (89.7%) (p<0.0001).

Bernardi et al., 2016

This is a prospective double-reading study in which 9,672 asymptomatic women over the age of 49 in Northern Italy underwent both DM and DBT. There were two reading strategies. In the first strategy, two independent, experienced breast radiologists sequentially interpreted the separately acquired DM images and then the combined separately acquired DM+DBT images. In the second strategy, two independent, experienced breast radiologists sequentially interpreted synthetic DM images derived from DBT acquisition and then the combined synthetic DM+DBT images. A screening was deemed positive and recall initiated if either reader interpreted either one of the sequential images as positive. Screening detected 90 cancers in 85 women; 76 of the cancers detected were invasive. Of the 76 invasive cancers, 46 were detected by both standard DM alone and by the integrated screenings (DM+DBT and synthetic DM+DBT), and 28 were detected only by the integrated screening.

The cancer detection rate was 6.3 per 1,000 (95% CI 4.8 to 8.1) for DM alone compared with 8.5 per 1,000 (95% CI 6.7 to 10.5) for integrated DM+DBT and 8.8 per 1,000 (95% CI 7.0-10.8) for integrated synthetic DM+DBT. The improvements in the cancer detection rate for integrated strategies over DM alone were more pronounced in women under the age of 60 and women with heterogeneously or extremely dense breasts.

Compared to both of the integrated strategies, DM alone had a lower false-positive recall rate (3.42% vs. 3.97% for DM+DBT and 4.45% for synthetic DM+DBT.) The lower false-positive recall rate for DM alone was most pronounced among women with heterogeneously or extremely dense breasts.

The trial did not follow women in order to determine the rate of interval cancers that were not detected by screening.

Houssami & Turner, 2016

This is a rapid review and meta-analysis of cancer detection and recall rates for DBT in women with dense breasts. The authors divided the trials into prospective studies that compared screening detection in the same subjects between DM and DBT (Ciatto, 2013; Lang 2016; Bernardi, 2016; Tagliafico 2016), and retrospective studies that compared screening detection in different groups of subjects (Rose, 2013; McCarthy, 2014; Conant, 2016; Rafferty 2016). It should be noted that in one of the included trials (Tagliafico, 2016), the patients had been referred for adjunctive screening after a negative digital mammogram. In the meta-analysis of prospective studies, the incremental cancer detection rate was 3.9 additional cancers identified per 1,000 screens with DBT (95% CI 2.7 to 5.1). In the meta-analysis of the retrospective studies, the incremental cancer detection rate was 1.4 additional cancers identified per
1,000 screens with DBT (95% CI 0.9 to 2.0). Pooled estimates for the difference in recall rates could only be estimated from the retrospective trials; in that analysis, DBT resulted in 23.3 fewer recalls per 1,000 screens compared to DM (95% CI -29.9 to -16.8).

Summary of Additional Studies

Among the additional studies submitted during public comment or identified through searches:

- Five found increases in the cancer detection rate with DBT compared to DM (Sharpe, 2016; Rose, 2014; Lang, 2016; Conant, 2016; Bernardi, 2016), whereas one found no statistically significant difference (McDonald, 2016). However, in one of these trials (Sharpe, 2016), there was no statistically significant difference in the rate of invasive cancer detection, and the rate of recall did not differ between DBT and DM after accounting for other factors that influence recall in the mixed-effects logistic regression model.
- Four found decreases in the recall rate for DBT compared to DM (Sharpe, 2016; Rose, 2014; McDonald, 2016; Conant 2016), one study found an increase in the recall rate (Lang, 2016), and one found an increase in the false-positive recall rate (Bernardi, 2016).
- A rapid review and meta-analysis of trials with subgroups of women with dense breasts found an increase in the cancer detection rate and a decrease in the recall rate with DBT compared to DM (Houssami & Turner, 2016).

Economic Analyses

Kalra et al., 2016

This is a cost-effectiveness analysis of annual screening using DBT from the federal payer perspective over a lifetime horizon. The clinical inputs were derived from Friedewald et al.’s 2014 study that showed a 1.2 per 1,000 screenings increase in cancer detection rate and 16 per 1,000 screenings decrease in the recall rate with DBT over DM. Costs were determined from the 2015 Medicare reimbursement values and the costs, utilities, and disutility associated with invasive and noninvasive cancers were derived from published literature. There authors noted that there is no reference standard for the disutility associated with false-positive recall. The standard 3% annual discounting was applied. Across all age groups in the base-case scenario, DBT resulted in a net gain of 0.04 quality-adjusted life-years (QALYs) over DM; the cost per QALY gained over DM was estimated at $20,300. In the probabilistic sensitivity analysis (in which inputs are varied to create 10,000 simulations), DBT was cost-effective compared to DM at a willingness-to-pay threshold (WTP) of $100,000 per QALY in 63.2% of the scenarios. In the deterministic sensitivity analyses, DBT remains cost-effective (at a WTP of $100,000 per QALY) as long as the recall rate is reduced by at least 1 per 1,000 screens and the cost of DBT does not exceed $250 more than the cost of DM alone. One limitation of this analysis was “an inability to capture the downstream costs of work-up for false positive cases” and the authors noted that their model accounted only for radiologic biopsies, not costlier surgical biopsies. Additionally, the clinical inputs were derived from large academic practices with uncertain congruence to community practices.

Bonafede et al., 2015

This is an economic modeling study designed to estimate the cost impact of full conversion from DM to DM+DBT in a hypothetical commercial managed care plan. The critical input pertaining to the use of
follow-up diagnostic services for people undergoing DM+DBT is based on unpublished proprietary market research data furnished by Truven Health Analytics. Additionally, the model assumes that DM+DBT would shift the distribution of diagnosed breast cancers toward earlier stages, an assumption that is not clearly supported in the literature (cf. Lang, 2016). In the base-case analysis, the conversion from DM to DBT+DM would save $0.20 per member per month (PMPM) with a range (depending on the assumed rate of follow-up services for DBT+DM) of $0.37 PMPM to $0.03 PMPM.

*Lee et al., 2014*

This is a cost-effectiveness analysis comparing the biennial screening with DBT+DM to biennial screening with DM alone among women aged 50 to 74 with dense breasts from a federal payer perspective over a lifetime horizon. The operating characteristics of DBT+DM compared with DM for all women (not exclusive to women with dense breasts) were derived from the Oslo screening trial (Skaane, 2013); those operating characteristics were assumed to be the best-case scenario and the base-case scenario relied on more modest estimates of the performance of DBT+DM. The authors assumed an additional cost of $50 for DBT+DM over the cost of DM alone, but because of uncertainty in that estimate they used a wide range of costs in the sensitivity analysis. In the base case analysis, DBT+DM resulted in a gain of 0.007 QALYs over DM with an incremental cost per QALY gained of $53,893. In the sensitivity analysis, the best-case scenario for DBT performance, the incremental cost per QALY gained was $26,107 and $792,264 in the worst case scenario for DBT performance. Assuming a WTP of $100,000 per QALY, DBT+DM remains cost-effective up to an added cost of $87 over the cost of DM. A probabilistic sensitivity analysis was not performed.

**Conclusions**

The evidence for using DBT for breast cancer screening is limited to observational studies, most of which have methodological limitations and inadequate follow-up periods. Thus, the effects of DBT on all-cause mortality, breast cancer morbidity, and breast cancer stage at diagnosis are unknown. Two studies with adequate follow-up to ascertain interval cancer rates and thereby permit calculation of sensitivity and specificity reached differing conclusions; one of these trials showed increased sensitivity and similar specificity with DBT+DM, and the other study showed identical sensitivity and improved specificity with DBT+DM. Some of the conflicting results may be accounted for by differences in reading models and recall patterns between the United States and Europe. There is low-quality evidence with mixed results that DBT+DM improves cancer detection rates. There is low-quality evidence that DBT+DM reduces recall rates and increases the positive predictive value of recall compared to DM alone, particularly when limited only to studies that were performed in the United States. Among the U.S.-based studies, reduction in recall rate is the most consistent finding (see Table 4). However, with the exception of the population of women with dense breasts, there are no meta-analytic estimates available for any of the outcomes.
POLICY LANDSCAPE

Quality measures

No quality measures related to DBT specifically were identified when searching the National Quality Measures Clearinghouse.

Payer coverage policies

Coverage policies were assessed for Aetna, Cigna, Moda, and Regence. Aetna, Moda, and Regence do not cover DBT because of insufficient evidence for its effectiveness. Although Cigna did not cover DBT for routine breast cancer screening under its previous policy, Cigna revised its coverage policy on August 23, 2016, to permit DBT for routine breast cancer screening based on recent guidance from the National Comprehensive Cancer Network (Cigna, 2016).

The Washington Medicaid program covers DBT when performed with a screening mammography for patients aged 40 to 74 who are candidates for screening mammography. Prior authorization is required for mammograms with or without DBT for patients age 39 or younger.

For Medicare, no National Coverage Determinations or Local Coverage Determinations related to DBT were identified.

Professional society guidelines

The U.S. Preventive Services Task Force issued an I recommendation for DBT in 2016, concluding that there was insufficient evidence to assess the benefits and harms of DBT for screening (USPSTF, 2016), based on the AHRQ systematic review (Melnikow et al., 2016a). Furthermore, the USPSTF also issued an I recommendation for adjunctive or supplemental screening, including DBT, for women with dense breasts.

Similarly, the American Congress of Obstetricians and Gynecologists (ACOG), American Cancer Society (ACS), American College of Physicians (ACP), and American Academy of Family Physicians (AAFP) all considered DBT in their breast cancer screening guidelines, but concluded that current evidence is insufficient to assess its effectiveness (ACOG, 2011; Oeffinger et al., 2015; Wilt, Harris, & Qaseem, 2015; AAFP, 2016).

Although the American College of Radiology (ACR) did not address DBT in its previous 2010 breast cancer screening recommendations (Lee et al., 2010), it released a position statement in November 2014, which states that DBT is no longer investigational and has demonstrated improvement in outcomes compared to digital mammography. The ACR (n.d.) summarizes its own position statement as follows:

“The ACR position on DBT is that it is no longer investigational. Tomosynthesis has been shown to improve key screening parameters compared to digital mammography.” The College applauds the decision by the Centers for Medicare and Medicaid Services (CMS) to facilitate access to these exams by covering beneficiaries for tomosynthesis and urges private payers to do the same.
Under its recommendation for an annual screening mammogram for average risk women aged 40 and over, the National Comprehensive Cancer Network (NCCN) recently added, “Consider tomosynthesis” (p. 7, 2016).

REFERENCES

Evidence Sources


**Other Citations**


Coverage guidance is prepared by the Health Evidence Review Commission (HERC), HERC staff, and subcommittee members. The evidence summary is prepared by the Center for Evidence-based Policy at Oregon Health & Science University (the Center). This document is intended to guide public and private purchasers in Oregon in making informed decisions about health care services.

The Center is not engaged in rendering any clinical, legal, business or other professional advice. The statements in this document do not represent official policy positions of the Center. Researchers involved in preparing this document have no affiliations or financial involvement that conflict with material presented in this document.

### APPENDIX A. GRADE INFORMED FRAMEWORK – ELEMENT DESCRIPTIONS

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balance of benefits and harms</td>
<td>The larger the difference between the desirable and undesirable effects, the higher the likelihood that a strong recommendation is warranted. An estimate that is not statistically significant or has a confidence interval crossing a predetermined clinical decision threshold will be downgraded.</td>
</tr>
<tr>
<td>Quality of evidence</td>
<td>The higher the quality of evidence, the higher the likelihood that a strong recommendation is warranted.</td>
</tr>
<tr>
<td>Resource allocation</td>
<td>The higher the costs of an intervention—that is, the greater the resources consumed in the absence of likely cost offsets—the lower the likelihood that a strong recommendation is warranted.</td>
</tr>
<tr>
<td>Values and preferences</td>
<td>The more values and preferences vary, or the greater the uncertainty in values and preferences, the higher the likelihood that a weak recommendation is warranted.</td>
</tr>
<tr>
<td>Other considerations</td>
<td>Other considerations include issues about the implementation and operationalization of the technology or intervention in health systems and practices within Oregon.</td>
</tr>
</tbody>
</table>

### Strong recommendation

**In Favor:** The subcommittee concludes that the desirable effects of adherence to a recommendation outweigh the undesirable effects, considering the balance of benefits and harms, resource allocation, values and preferences, and other factors.

**Against:** The subcommittee concludes that the undesirable effects of adherence to a recommendation outweigh the desirable effects, considering the balance of benefits and harms, resource allocation, values and preferences, and other factors.

### Weak recommendation

**In Favor:** The subcommittee concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, considering the balance of benefits and harms, resource allocation, values and preferences, and other factors, but further research or additional information could lead to a different conclusion.

**Against:** The subcommittee concludes that the undesirable effects of adherence to a recommendation probably outweigh the desirable effects, considering the balance of benefits and harms, cost and resource allocation, and values and preferences, but further research or additional information could lead to a different conclusion.

### Confidence in estimate rating across studies for the intervention/outcome

**High:** The subcommittee is very confident that the true effect lies close to that of the estimate of the effect. Typical sets of studies are RCTs with few or no limitations and the estimate of effect is likely stable.

**Moderate:** The subcommittee is moderately confident in the estimate of effect: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. Typical sets of

---

4 Includes risk of bias, precision, directness, consistency, and publication bias.
Digital Breast Tomosynthesis (3D Mammography) for Breast Cancer Screening in Average Risk Women
Approved 3/9/2017

studies are RCTs with some limitations or well-performed nonrandomized studies with additional strengths that guard against potential bias and have large estimates of effects.

Low: The subcommittee’s confidence in the estimate of effect is limited: The true effect may be substantially different from the estimate of the effect. Typical sets of studies are RCTs with serious limitations or nonrandomized studies without special strengths.

Very low: The subcommittee has very little confidence in the estimate of effect: The true effect is likely to be substantially different from the estimate of effect. Typical sets of studies are nonrandomized studies with serious limitations or inconsistent results across studies.
# Appendix B. GRADE Evidence Profile

<table>
<thead>
<tr>
<th>Quality Assessment (Confidence in Estimate of Effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Studies</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td><strong>All-cause mortality</strong></td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td><strong>Breast cancer morbidity</strong></td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td><strong>Test performance characteristics/Cancer detection rate (CDR)</strong></td>
</tr>
<tr>
<td>17 (for CDR)</td>
</tr>
<tr>
<td>2 (for sensitivity and specificity)</td>
</tr>
<tr>
<td><strong>Stage at diagnosis</strong></td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td><strong>Recall rate/False-positive rate</strong></td>
</tr>
<tr>
<td>16</td>
</tr>
</tbody>
</table>

Digital Breast Tomosynthesis (3D Mammography) for Breast Cancer Screening in Average Risk Women
Approved 3/9/2017
APPENDIX C. METHODS

Scope Statement

Populations

Women between the ages of 40 and 74 years referred for breast cancer screening

Population scoping notes: Excludes women with a personal history of breast cancer, clinically significant BRCA gene mutations, Li-Fraumeni syndrome, Cowden syndrome, hereditary diffuse gastric cancer or other familial breast cancer syndromes, high-risk lesions (ductal or lobular carcinoma in situ, atypical ductal or lobular hyperplasia), or previous large doses of chest radiation therapy (≥ 20 Gy) before age 30 years.

Interventions

DBT (3-D mammography) in conjunction with standard 2-D digital mammography

Intervention exclusions: None

Comparators

Standard 2-D mammography with or without computer-aided diagnosis

Considered but not selected: No screening, MRI, ultrasound

Outcomes

Critical: All-cause mortality, breast cancer morbidity

Important: Test performance characteristics, cancer stage at diagnosis, recall rate/false-positive test results

Considered but not selected for the GRADE table: Cancer-specific mortality, radiation exposure, PPV for recalls, PPV for biopsies

Key Questions

KQ1: What is the comparative effectiveness of digital breast DBT as a primary screening modality in women referred for breast cancer screening?

KQ2: Does the comparative effectiveness of DBT vary by the following characteristics:
   a. Age
   b. Race or ethnicity
   c. Breast density

KQ3: In a screening population, how do the test characteristics of 3-D/2-D mammography compare to those of standard 2-D mammography?

KQ4: What are the harms of 3-D/2-D mammography compared to standard 2-D mammography alone?

KQ5: If DBT is used as a primary screening modality, what is the optimal screening interval, and does that interval vary according to the characteristics listed in Key Question 2?
**Search Strategy**
A full search of the core sources was conducted to identify systematic reviews, meta-analyses, technology assessments, and clinical practice guidelines using the terms Digital Breast Tomosynthesis and 3-dimensional (3-D) mammography. Searches of core sources were limited to citations published in the past five years.

The core sources searched included:
- Agency for Healthcare Research and Quality (AHRQ)
- Blue Cross/Blue Shield Health Technology Assessment (HTA) program
- BMJ Clinical Evidence
- Canadian Agency for Drugs and Technologies in Health (CADTH)
- Cochrane Library (Wiley Interscience)
- Hayes, Inc.
- Institute for Clinical and Economic Review (ICER)
- Medicaid Evidence-based Decisions Project (MED)
- National Institute for Health and Care Excellence (NICE)
- Tufts Cost-effectiveness Analysis Registry
- Veterans Administration Evidence-based Synthesis Program (ESP)
- Washington State Health Technology Assessment Program

A MEDLINE® search was conducted to identify systematic reviews, meta-analyses, and technology assessments published in the past five years. A MEDLINE® search was then conducted to identify randomized control trials and cohort studies that would have been included in the identified systematic reviews except for being published after the search dates of the systematic reviews.

Searches for clinical practice guidelines were limited to those published since 2011. A search for relevant clinical practice guidelines was also conducted, using the following sources:
- Australian Government National Health and Medical Research Council (NHMRC)
- Centers for Disease Control and Prevention (CDC) – Community Preventive Services
- Choosing Wisely
- Institute for Clinical Systems Improvement (ICSI)
- National Guidelines Clearinghouse
- New Zealand Guidelines Group
- NICE
- Scottish Intercollegiate Guidelines Network (SIGN)
- United States Preventive Services Task Force (USPSTF)
- Veterans Administration/Department of Defense (VA/DOD)

**Inclusion/Exclusion Criteria**
Studies were excluded if they were not published in English, did not address the scope statement, or were study designs other than systematic reviews, meta-analyses, technology assessments, or clinical practice guidelines.
APPENDIX D. APPLICABLE CODES

<table>
<thead>
<tr>
<th>CODES</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICD-10 Diagnosis Codes</strong></td>
<td></td>
</tr>
<tr>
<td>Z12.31</td>
<td>Encounter for screening mammogram for malignant neoplasm of breast</td>
</tr>
<tr>
<td><strong>CPT Codes</strong></td>
<td></td>
</tr>
<tr>
<td>77051</td>
<td>Computer-aided detection with further physician review for interpretation; diagnostic mammography</td>
</tr>
<tr>
<td>77052</td>
<td>Computer-aided detection with further physician review for interpretation; screening mammography</td>
</tr>
<tr>
<td>77055</td>
<td>Mammography; unilateral</td>
</tr>
<tr>
<td>77056</td>
<td>Mammography; bilateral</td>
</tr>
<tr>
<td>77057</td>
<td>Screening mammography; bilateral</td>
</tr>
<tr>
<td>77061</td>
<td>Digital breast tomosynthesis; unilateral</td>
</tr>
<tr>
<td>77062</td>
<td>Digital breast tomosynthesis; bilateral</td>
</tr>
<tr>
<td>77063</td>
<td>Screening digital breast tomosynthesis; bilateral (in addition to primary screening mammography procedure)</td>
</tr>
<tr>
<td><strong>HCPCS Level II Codes</strong></td>
<td></td>
</tr>
<tr>
<td>G0202</td>
<td>Screening mammography; producing direct digital image, bilateral, all views</td>
</tr>
<tr>
<td>G0204</td>
<td>Diagnostic mammography, producing direct digital image, bilateral, all views</td>
</tr>
<tr>
<td>G0206</td>
<td>Diagnostic mammography, producing direct digital image, unilateral, all views</td>
</tr>
<tr>
<td>G0279</td>
<td>Diagnostic digital breast tomosynthesis; unilateral or bilateral</td>
</tr>
</tbody>
</table>

Note: Inclusion on this list does not guarantee coverage.