# New York Department of Health

## Dossier Summary and Response

**Topic**: Breast Tomosynthesis **Date**: March 25, 2016

### Dossier Submission

Hologic, Inc. submitted a dossier on breast tomosynthesis on December 9, 2015. The dossier was completed in accordance with the Department's instructions and included 26 articles for review published between 2013 and 2015. Of the submitted articles, 15 were rated by the submitter as having good methodologic quality and 11 were rated as being of fair methodologic quality. The submitted articles provided comparative outcomes on three-dimensional (3D) mammography or digital breast tomosynthesis (DBT), with or without two-dimensional (2D) digital mammography (DM), in comparison to 2D DM alone.

### Dossier Review Process

The Center for Evidence-based Policy (Center) provided a review of the submitted dossier. Submitted articles were independently assessed for inclusion, methodological quality, and reported results. Literature searches of the MEDLINE® (Ovid) database and the Center's core sources[1] (a select group of resources considered high quality due to being independent and using systematic methods) were conducted to identify any additional relevant evidence.

### Review Results

Evidence Evaluation – Included Studies

Center staff performed a search of its core sources and a MEDLINE® (Ovid) search to identify any additional articles relevant to the topic. The search methodology is detailed in Appendix A. When reviewing the studies either submitted with the dossier or identified by the subsequent search, only comparative studies were considered for evaluation of efficacy. Based on the dossier submitter's inclusion criteria, only studies that included an asymptomatic screening population were included. Included studies were limited to English language, systematic reviews (SRs) with or without meta-analyses (MAs), randomized controlled trials (RCTs), or observational studies. Case series were additionally considered to evaluate harms. In addition, only patient important outcomes have relevance for New York Department of Health. The rationale for study inclusion can be found in the New York Department of Health Dossier

[1] Center core sources searched include Hayes, Inc., Cochrane Library (Wiley Interscience), the United Kingdom National Institute for Health and Care Excellence (NICE), the Blue Cross/Blue Shield Health Technology Assessment (HTA) program, the Veterans Administration Technology Assessment Program (VATAP), *BMJ Clinical Evidence*, the Canadian Agency for Drugs and Technologies in Health (CADTH), the Washington State Health Technology Assessment Program, the United States Preventive Services Task Force (USPSTF), and the Agency for Healthcare Research and Quality (AHRQ).

Methods Guidance (New York Department of Health, 2015). Exclusion criteria were selected prior to review of the studies, and study methods were assessed prior to review of outcomes to eliminate bias.

Exclusion criteria included:

- Population with breast abnormalities, dense breasts, or at high risk for breast cancer
- Non-comparative studies
- Historically-controlled cohort studies
- Duplicate information from a research study published in more than one source once (only the highest quality, most recent publication with outcome of interest was included)
- Systematic reviews that included only studies that were summarized by more comprehensive SRs or SRs of higher quality and/or that were more recently published
- Studies identified that were included in a summarized SR or technology assessment (TA)

Follow-up of 12 months or greater is needed to detect interval cancers (cancers that were not detected by screening) in order to calculate the sensitivity[2], specificity[3], and negative predictive value[4] of the test. If a study does not include a follow-up period, the calculated sensitivity will always be 100% since women who screened negative, but actually had cancer at the time of screening, would not be identified. The Agency for Health Care Research and Quality (AHRQ), in a recent SR (Melnikow et al., 2016), used a 12-month follow-up period, with repeat imaging at one year, as a necessary criterion for study inclusion. The dossier submitter did not specify sensitivity, specificity, and negative predictive value as outcomes of interest, thus study follow-up length was not included in the Center's inclusion/exclusion criteria. However, sensitivity, specificity, and negative predictive value are important test characteristics, and there is a lack of information on these outcomes for DBT in this dossier submission and review.

Center staff identified 11 recent SRs comparing DBT to DM, however, only three SRs met the specified inclusion and exclusion criteria, one of which was included in the submitted dossier (Washington Health Technology Assessment Program [WA HTA], 2014). The Medline® (Ovid) database search did not identify any additional studies to those provided by the submitter. The search strategy and list of studies reviewed in full with reason for exclusion are included in Appendices A and B, respectively.

[2] The number of positive tests among women who have breast cancer divided by the total number of women with breast cancer in the population

[3] The number of negative tests among women who do not have breast cancer divided by the total number of women without breast cancer in the population

[4] The number of truly negative images divided by the total number of negative images (the likelihood that a women who has a negative test does not have cancer)

Review of the included dossier materials resulted in exclusion of 11 of the 26 submitted articles based on study design, population, intervention, or comparator (see Table 3 for further description of studies and exclusion criteria). Ten of the submitted articles (nine studies) were included in the summarized SRs, and thus not assessed for methodologic quality by Center staff. Individual publications of these nine studies were reviewed only to clarify information reported in the SRs, when needed (Ciatto et al., 2013; Destounis et al., 2014; Friedewald et al., 2014; Greenberg et al., 2014; Haas et al., 2013; Houssami et al., 2014; McCarthy et al., 2014; Rose et al., 2013; Skaane et al., 2013a; Skaane et al., 2013b). Table 1 includes a complete list of included articles, and associated methodological quality ratings, sample size, and findings identified in the searches described above. Study methodologic quality was rated by Center using the same quality assessment forms as provided by the submitter. Appendix D includes the submitter's and Center's assessments for all included studies.

## Evidence Review

This section provides an overview of included studies and a summary of the findings regarding effectiveness, harms and costs related to DBT. The quality ratings included in this section refer to the ratings by the Center unless otherwise specified. Table 1 provides a further summary of the studies with more detail than included in the summary below.

### Overview of Included Studies

Three SRs are included in this review (Melnikow et al., 2016; Nelson et al., 2016; WA HTA, 2014). All of the SRs were rated as having good methodological quality. The Melnikow (2016) is recently published SR from AHRQ on the performance characteristics of DBT either alone or in combination with 2D mammography compared to 2D mammography as a primary screening test for breast cancer. Authors performed an extensive literature search from January 2000 to October 2015 with the following inclusion criteria: 1) conducted in screening populations; and 2) test characteristics evaluated with a comprehensive reference standard applied to negative and positive tests. Authors identified one single prospective cohort study (STORM) that included 7,292 women aged 48 years or older in Northern Italy (Ciatto et al., 2013; Houssami et al., 2014). The AHRQ report summarized the outcomes for eight additional prospective and retrospective cohort studies which did not meet their inclusion criteria (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).

In the STORM prospective cohort study (Ciatto et al., 2013; Houssami et al., 2014), radiologists read digital 2D mammograms sequentially and then read and interpreted the DBT images in combination with the 2D images and DBT in the same session. The study utilized double-reading, and screening participants were recalled if there was a positive read from either reader on the 2D or 3D images. Median follow-up was 19.7 months. The AHRQ report (Melnikow et al.,

2016) rated this study as good quality because it included an asymptomatic screening population, a reference standard was applied to positive and negative results, and follow-up was more than one year.

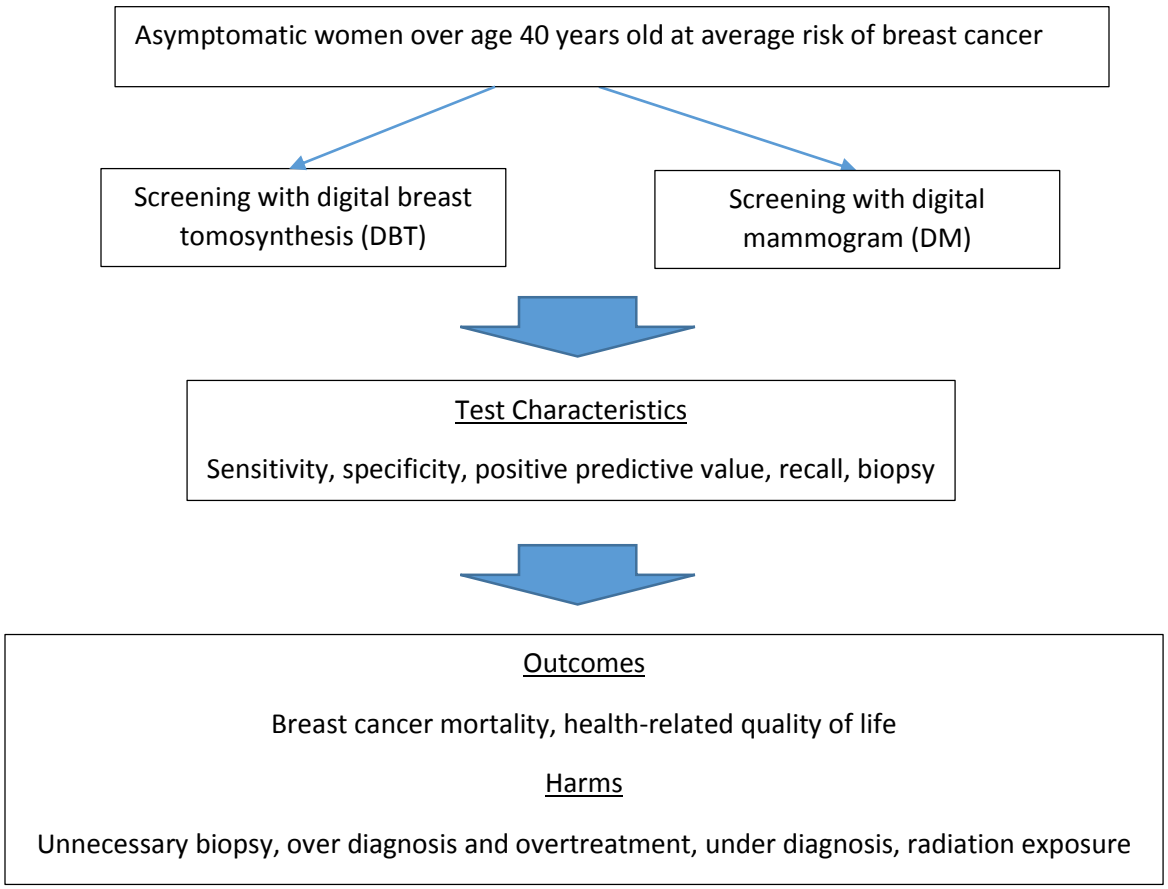
Nelson and colleagues (2016), in a SR completed to inform the update of the 2009 United States Preventive Services Taskforce Recommendation, evaluated the harms of mammography, including an assessment of DBT compared to DM. The authors used an extensive search strategy performed through December 2014. In addition to the STORM prospective cohort described above (Ciatto et al., 2013), Nelson and colleagues (2016) included four additional cohort studies that were performed in asymptomatic screening populations (Friedewald et al., 2014; Haas et al., 2013; Rose et al., 2013; Skaane et al., 2013b). The authors rated the overall quality of the evidence as poor.

In the largest U.S. cohort included by Nelson and colleagues (2016), Friedewald and colleagues (2014) retrospectively analyzed 454,850 images before and after the introduction of DBT at 13 medical centers. This study was performed on a population level, and data were limited to what was reported in the screening records. Haas and colleagues (2013) performed a retrospective contemporaneous analysis of DBT with DM compared to mammography alone (n=13,158) at four clinical sites in Connecticut. In the retrospective cohort study by Rose and colleagues (2013), authors used a pre-post design to evaluate the performance of DBT with DM at a multisite community-based breast center in Texas compared to DM alone at the same multisite center one year prior (n=13,856).

In the remaining study, Skaane and colleagues (2013b) conducted a prospective cohort of 12,621 women in Norway who received DBT with DM or mammography alone. Each image was read and scored separately by two readers. Cases that had one or more images with an elevated score were referred to an arbitration process, where two or more radiologists reviewed the imaging results and conferred on their assessment. Patients were only recalled after the arbitration process. This practice is dissimilar to the standard practice in the U.S., where one radiologist reviews an image and determines necessary follow-up. A major limitation of the five articles included in the Nelson (2016) SR is that none of the studies followed up with patients after the initial screening to detect interval cancers, and therefore the sensitivity of imaging is overestimated.

The WA HTA (2014) prepared a good quality SR addressing the question of effectiveness and harms of DBT compared to DM in a screening population for the WA HTA. A search using the analytic framework below (Figure 1) was performed for studies published from January 1990 to November 2014.

Figure 1. *Analytic Framework*




Nine studies in 10 articles (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) were included in the WA HTA SR (2014) and were all rated as having poor methodologic quality by the review authors. The WA HTA (2014) review used the QUADAS-2[5] domains for diagnostic accuracy studies including patient selection, index testing, reference standard, and timing within study. Only one study had a follow-up period of one year or more, and there was a 20% drop-out rate in that study (Destounis et al., 2014). The WA HTA (2014) SR included four U.S. retrospective cohort studies not discussed previously. In the New York-placed retrospective cohort study by Destounis and colleagues (2014), 524 women choosing to undergo DBT plus DM were compared to women who underwent DM alone during the same time period. In the retrospective cohort study by Greenberg and colleagues (2014) , 20,943 women volunteering to undergo DBT in addition to mammography were compared to 38,674 mammograms occurring during the same time period

[5] The QUADAS tool is used for rating the quality of diagnostic accuracy studies and was developed in 2003 in a collaboration funded by the United Kingdom Health Technology Assessment program. It is recommended for use by the National Institute for Health and Clinical Excellence (NICE) and AHRQ.

in Washington, D.C. Women were asked to pay $50 for the DBT, but were offered a free DBT exam if they could not afford the cost. The newer DBT images were studied immediately after implementation at the study sites, leaving no adjustment period for learning curves. The U.S.-based retrospective analyses by Lourenco and colleagues (2014) and McCarthy and colleauges (2014) used a pre-post design to compare DBT plus DM exams to DM exams performed the year prior at the same study sites. Additional study details are included in Table 2.

Two additional studies that were not included in the SRs/TAs described above were included in the dossier submission. Lang and colleagues (2015) performed a poor quality prospective analysis of women receiving either DBT alone or DM alone. In this Swedish cohort of 7,500 women, an arbitration process was performed prior to patient recall. There was no follow-up to detect interval cancers. The second study is a U.S.-based retrospective cohort comparing DBT plus DM (n=8591) to DM alone (n=9364) performed contemporaneously (Durand et al., 2014). Digital mammography was available at outpatient clinics and mobile imaging site, while DBT was only available at a tertiary care hospital and one of the outpatient sites mid-way through the study. There were baseline differences in age and breast cancer risk factors between groups, and there was no follow-up to detect interval cancers. Table 1 provides additional information on the SRs/TAs and individual studies included in this dossier review.

Effectiveness

None of the identified studies addressed the clinically-important outcomes of breast-cancer related morbidity, breast cancer recurrence or second breast cancers, or mortality. The effectiveness outcomes considered in this dossier evaluate the diagnostic accuracy of DBT, and the outcomes selected reflect the current incomplete state of evidence on the diagnostic accuracy of DBT. Only one prospective cohort study had a follow-up great than 12 months and reference testing to assess interval cancers which enables one to estimate the sensitivity, specificity, and negative predictive value of the test. The AHRQ (Melnikow et al., 2016) review, which reported outcomes from the STORM prospective study (Ciatto et al., 2013; Houssami et al., 2014), reported the sensitivity of breast cancer detection for a single read of DBT combined with 2D mammography was 0.85 (95% confidence interval [CI], 0.74 to 0.92) compared to 0.54 (95% CI, 0.42 to 0.65) for 2D mammography. Specificity of breast cancer detection for DBT plus 2D mammography was 0.97 (95% CI, 0.96 to 0.98) compared to 0.96 (95% CI, 0.95 to 0.98) for 2D DM alone.

Table 1. *Evidence Review – Included References*

| Citation, Study Details | Dossier QA | Center QA | # of Studies (k) / Population (n) | Study Summary and Findings | Comments[6] |
|---|---|---|---|---|---|
| **Systematic Reviews** | | | | | |
| AHRQ (2016)<br><br>Search Dates<br>January 2000 to October 2015<br><br>Included Study Designs<br>Prospective cohort | *Not included* | Good | k = 1<br>total n = 7,292<br><br>*SR's quality assessment of individual studies:* Good | Comparators<br>DBT + DM vs. DM<br><br>Outcomes<br>*Sensitivity for Breast Cancer:* 0.85 (95% CI, 0.74 to 0.92) vs. 0.54 (95% CI, 0.42 to 0.65)<br>*Specificity for Breast Cancer:* 0.97 (95% CI, 0.96 to 0.98) vs 0.96 (95% CI, 0.95 to 0.97) | Included studies<br>Ciatto et al. (2013), Houssami et al. (2014)<br><br>Summarized in evidence tables[7]<br>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) |
| Nelson et al. (2016)<br><br>Search Dates<br>Through December 2014<br><br>Included Study Designs | *Not included* | Good | k=5<br>total n = 517,011<br><br>*SR's quality assessment of individual studies:* Poor | Comparators<br>DBT + DM vs. DM<br><br>Outcomes<br>*Recall Rate*: Significantly lower for DBT+ DM vs. DM across studies<br>One U.S. study reported 16 less recalls per 1000 screens (p<0.001) (Friedewald et al., 2014) | Included studies<br>Ciatto et al. (2013), Friedewald et al. (2014), Haas et al. (2013), Rose et al. (2013), Skaane et al. (2013a)<br><br>Evidence limited by lack of RCTs, comparability of results not reported, and outcomes not reported uniformly |

[6] Included studies in bold-face type were submitted in the dossier

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

| Citation, Study Details | Dossier QA | Center QA | # of Studies (k) / Population (n) | Study Summary and Findings | Comments[6] |
|---|---|---|---|---|---|
| SRs, RCTs, observational studies | | | | *Biopsy Rate*: Increase of 1.3 biopsies per 1,000 screens for DBT+ DM compared to DM (p<0.001) (Friedewald et al., 2014) | |
| WA HTA (2014)<br><br><u>Search Dates</u><br>January 1990-November 2014<br><br><u>Included Study Designs</u><br>Observational studies | Good | Good | k = 9<br>total n = 313,298<br><br>*SR's quality assessment of individual studies:* Poor | <u>Comparators</u><br>DBT+ DM vs. DM<br><br><u>Outcomes*</u><br>*Cancer Detection Rate (CDR):* 4 to 6 / 1,000 vs. 3 to 5 / 1,000<br>*Recall Rate:* 80 to 140 / 1,000 vs. 100 to 160 / 1,000<br>*Biopsy Rate:* 4 to 6 / 1,000 vs. 3 to 5 / 1,000<br>*PPV Biopsy:* 25 to 30% vs. 20 to 25%<br><br>*Meta-analysis not performed for outcomes, significance not reported | <u>Included studies</u><br>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)<br><br>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% drop-out rate in the study with 12 month follow-up (Destounis et al., 2014)<br><br>Some of the studies had possible selection bias<br><br>Authors reported a moderate to high degree of uncertainty in recall rate, biopsy rate, and CDR<br><br>There is a low to moderate degree of uncertainty for the PPV of biopsy |

| Citation, Study Details | Dossier QA | Center QA | # of Studies (k) / Population (n) | Study Summary and Findings | Comments[6] |
|---|---|---|---|---|---|
| **Diagnostic Accuracy Studies** | | | | | |
| Lang et al. (2015)<br><br>Design<br>Prospective cohort (1 arm)<br><br>Location<br>Sweden<br><br>Test<br>Mammomat Inspiration, Siemens AG | Good | Poor | n = 7,500 | Comparators<br>DBT only vs. DM only<br><br>Outcomes<br>*CDR:* 8.9 / 1,000 vs. 6.3 / 1,000 (p<0.0001)<br>*Percent of cancers that were invasive:* 85% vs. 89%<br>*Recall Rate:* 3.8 % vs. 2.6% ( p<0.0001)<br>*Biopsy Rate:* NR<br>*PPV Recall:* 24% vs. 24%<br>*PPV Biopsy:* NR | No follow-up, interval cancers not detected and cancer detection rates likely over-estimated<br><br>Arbitration process prior to recall<br><br>Funded by Siemans AG |
| Durand et al. (2014)<br><br>Design<br>Retrospective cohort (2 arms)<br>Historical cohort also used<br><br>Location<br>U.S. | Good | Poor | n (DBT+ DM) = 8,591<br>n (DM) = 9,364 | Comparators<br>DBT+ DM vs. DM<br><br>Outcomes<br>*CDR:* No sig. difference<br>*Recall Rate:* Sig. lower for DBT+ DM for asymmetries and calcifications, but not masses or architectural distortions<br>*Biopsy Rate:* NR<br>*PPV Recall:* NR<br>*PPV Biopsy:* NR | No follow-up, interval cancers not detected and cancer detection rates likely over-estimated<br><br>Possibility for selection bias as DBT only available at tertiary care hospital and one of the outpatient clinics midway through the study<br><br>Age, breast density, and breast cancer risk factors were statistically significantly different between groups |

| Citation, Study Details | Dossier QA | Center QA | # of Studies (k) / Population (n) | Study Summary and Findings | Comments[6] |
|---|---|---|---|---|---|
| <u>Test</u><br>Mammogram: Selenia, Hologic<br>Tomosynthesis: Dimensions, Hologic | | | | | |

Abbreviations: CDR= cancer detection rate, DBT= digital breast tomosynthesis, DM= digital mammography, NR= not reported, PPV= positive predictive value, QA= quality assessment, RCT= randomized controlled trial, sig = significance, SR= systematic review

*Outcome #1: Cancer Detection Rate*

The AHRQ (Melnikow et al., 2016) SR included the reported cancer detection rate (CDR) from Houssami and colleagues (2014) in an evidence table, and also included in the evidence table a summary of the eight studies that were not assessed for methodologic quality or summarized in narrative form. Cancer detection rate was not reported in the Nelson (2016) SR. Six of the nine studies reviewed in the WA HTA (2014) review (Ciatto et al., 2013; Destounis et al., 2014; Friedewald et al., 2014; Haas et al., 2013; Rose et al., 2013; Skaane et al., 2013a) concluded that DBT significantly increases the CDR compared to DM. Results were consistent for women with dense breasts in studies that performed subgroup analysis based on breast density. Across these European and U.S. studies of asymptomatic women presenting for routine screening, the CDR for DBT was four to six cancers per 1,000 individuals, compared to three to five cancers per 1,000 individuals for DM. Authors of the WA HTA (2014) review reported having a moderate to high degree of uncertainty in this estimate due to study limitations. Most notably, eight of the nine studies did not follow patients to detect interval cancers, which would overestimate sensitivity. Additionally, some of the studies had baseline differences in which women in the DBT groups would have a higher risk of cancer, and therefore the CDR in this group may be elevated due to baseline differences (Destounis et al., 2014; Haas et al., 2013). Furthermore, several of the studies had a pre-post design, which limits comparison between the two populations as temporal differences may have influences differences in outcomes. The findings and limitations of the individual studies from the SRs/TAs are included in Table 2.

In a poor quality Swedish prospective cohort study in women aged 40 to 76 years, 7,500 women received one-view DBT and two-view DM on the same day. The images were interpreted separately by two radiologists who were blinded to the readings of the other one. If one or more of the readers interpreted an abnormality on one or both of the images, two or more readers re-evaluated the images and determined if that patient needed to be recalled. There was no follow-up for interval cancers. Cancer detection rates were higher among women receiving single-view DBT (8.9/1,000; 95% CI, 6.9 to 11.3) compared to two-view DM (6.3/1,000; 95%CI, 4.6 to 8.3) (Lang et al., 2015).

A U.S.-based poor quality retrospective cohort study compared 8,591 DBT plus 2D DM exams to 9,364 2D DM exams performed during the same time period at four clinical sites. Selection bias was possible as DBT was only available at the tertiary care hospital for the entire study period and one outpatient imaging clinic for half of the study period. The other two sites, which included an outpatient imaging clinic and a mobile breast imaging clinic offered only mammography. There were statistically differences in baseline characteristics of the populations. Women in the DBT plus 2D DM group were more likely to be younger, have a lower rate of positive family history, and were less likely to have dense breasts. These results would bias the study toward detecting cancers in the 2D mammography only group compared

to the DBT plus 2D mammography group. There was no follow-up in the study, making it impossible to detect interval cancers. In addition to comparing the two groups, study authors compared the results to a historic control group. The CDRs did not differ between DBT plus 2D mammography and 2D mammography alone. Additionally, CDRs did not differ between the DBT plus 2D mammography group and the historical control. More cancers were identified in the group of participants receiving 2D mammography alone compared to the historical group control (5.9 vs. 4.4 per 1,000 cancers) (Durand et al., 2014).

*Outcome #2: Recall Rate*

Recall rate was included in evidence tables of studies that were not assessed for methodologic quality nor described in the narrative of the AHRQ (Melnikow et al., 2016) SR. Nelson and colleagues (2016) reported significantly lower recall rate across studies (Ciatto et al., 2013; Friedewald et al., 2014; Haas et al., 2013; Rose et al., 2013; Skaane et al., 2013a). The Nelson review reported a lack of RCTs and that evidence is limited by lack of uniformity in populations and methods across studies. The six studies reviewed by the WA HTA (2014) reported significantly higher CDRs and significantly lower recall rates for DBT plus 2D mammography compared to mammography alone (Ciatto et al., 2013; Destounis et al., 2014; Friedewald et al., 2014; Haas et al., 2013; Rose et al., 2013; Skaane et al., 2013a) The recall rate was 80 to140 per 1,000 participants for DBT+ 2D DM compared to 100 to 160 per 1,000 participants in the 2D mammography only group. In two studies, recall rates were lower with DBT compared to DM for women with dense breasts (Haas et al., 2013; Skaane et al., 2013a) and similar across breast density categories in two other studies performing subgroup analysis (McCarthy et al., 2013; Rose et al., 2013). The WA HTA (2014) review reported a moderate to high degree of uncertainty in this estimate due to the study limitations. In addition to the limitations of these studies described above, one European study had an arbitration process for positive readings that is not similar to U.S. practices and which has been shown to lower the rate of recall (Skaane et al., 2013a). Table 2 includes a summary of the studies included in the SRs and TAs.

In the poor methodologic quality prospective cohort by Lang and colleagues (2015), the recall rate was higher among women receiving one-view DBT compared to two-view mammography (3.8% vs 2.6%, $p<0.0001$). This finding is different than most studies comparing DBT plus 2D mammography compared to 2D mammography alone, in which the recall rate is higher for 2D mammography only. Methodological differences were likely to have impacted this outcome. In the Lang and colleagues (2015) study, DBT was not read in conjunction with 2D DM. In addition, readers participated in an arbitration process prior to recall (Lang et al., 2015).

In the poor quality retrospective cohort of asymptomatic women receiving either DBT plus 2D mammography or 2D DM alone, there was a higher recall rate in the DBT plus 2D mammography group (Durand et al., 2014). However, overall recall rates are not reported.

Patients in the DBT plus 2D mammography group were less likely to be recalled for asymmetries (3.1% vs. 7.9%, p<0.0001). This group was also less likely to be recalled for calcifications (2.5% vs. 3.2%, p=0.0005). The recall rates for architectural distortion or masses were not different (2.5% for each group). Baseline differences in study populations may have introduced confounding factors (Durand et al., 2014).

*Outcome #3: Positive Predictive Value of Recall Leading to Confirmed Cancer*
Positive predictive value (PPV) of a positive test (recalled test) is defined as the percentage of recalls leading to a biopsy-confirmed cancer diagnosis. None of the SRs or TAs included in this dossier review compared this outcome in the different imaging modalities. For DBT alone, the WA HTA (2014) review reported that PPV of recall ranged from 4.6% to 18.8% across the nine studies. Details of the studies are included in Table 2.

The poor quality prospective cohort comparing one-view DBT to two-view DM concluded that the PPV of recall of both imaging modalities was 24% (Lang et al., 2015).

*Outcome #4: Positive Predictive Value of Biopsy Leading to Confirmed Cancer*
Positive predictive value of a biopsy is defined as the percentage of biopsies necessitated by imaging that led to a cancer diagnosis. The AHRQ (Melnikow et al., 2016) and Nelson (2016) reviews did not report this outcome. In the WA HTA (2014) review, authors reported a PPV of biopsy of 25% to 30% for DBT plus 2D mammography compared to 20% to 25% of 2D mammography alone. The review reported a low to moderate level of uncertainty in this estimate. Six individual studies summarized in the SRs/TAs reported PPV of biopsy leading to a cancer diagnosis (Destounis et al., 2014; Friedewald et al., 2014; Greenberg et al., 2014; Lourenco et al., 2014; McCarthy et al., 2014;  Rose et al., 2013), and in five of the six studies the PPV of biopsy was higher for the DBT plus DM group. This difference was reported to be statistically significant in only one study (Friedewald et al., 2014). The data are summarized in Table 2.

Table 2. *Summary of Studies Included in SRs/TAs*[8]

| Author (Year) Study Size Location | Dossier QA | Source QA | Study Design and Population Characteristics | CDR per 1,000 women (% invasive) | Recall Rate (%) | PPV Recall | PPV Biopsy | Comments |
|---|---|---|---|---|---|---|---|---|
| Ciatto (2013)<br>n = 7,292<br>Italy | Fair | Poor (WA HTA, 2014) | Prospective cohort ( 1 arm)<br>Population-based screening centers<br>Mean age: 58<br>Test: Selenia Dimensions, Hologic | DBT+ DM: 8.1<br>DM: 5.3<br>(p<0.0001) | DBT+ DM: 4.3%<br>DM: 5%<br>(NS) | NR | NR | No long-term follow-up; 1 abnormal read-flagged recall |
| Destounis (2014)<br>DBT = 524<br>DM = 524<br>New York | Fair | Poor (WA HTA, 2014) | Retrospective cohort (2 arm)<br>Community breast clinic<br>Mean age: 59<br>Test: Selenia Dimensions, Hologic<br>SecurView, Hologic | DBT+ DM: 5.4 (33%)<br>DM: 3.8 (50%)<br>(sig. NR) | DBT+DM: 4.2%<br>DM: 11.4%<br>(p<0.0001) | NR | DBT+ DM: 50.0%<br>DM: 16.7%<br>(sig. NR) | 1 year follow-up; 80% completion rate<br>Selection bias likely due baseline risk factors for breast cancer or abnormal imaging in the DBT group |
| Friedewald (2014) | Good | Poor (WA HTA, 2014) | Retrospective cohort (2 arm): Pre-post | DBT+ DM: 5.5 (75%)<br>DM: 4.3 (67%) | DBT+ DM: 8.9%<br>DM: 10.6% | DBT+ DM: 6.1%<br>DM: 4.1% | DBT+ DM: 29.2% | Insufficient follow-up<br>Pre-post design |

[8] Center staff abstracted information from the original study where information was not reported in an SR/TA or when there was conflicting information reported in SRs/TAs.

| Author (Year)<br>Study Size<br>Location | Dossier QA | Source QA | Study Design and Population Characteristics | CDR per 1,000 women (% invasive) | Recall Rate (%) | PPV Recall | PPV Biopsy | Comments |
|---|---|---|---|---|---|---|---|---|
| DBT+ DM exams = 173,663<br>DM exams = 281,187<br>U.S., Multi-state | | | 13 academic medical centers and breast diagnostic/screening centers<br>Mean age: 56.2 for DBT+ DM; 57.0 for DM<br>Test: Selenia Dimensions, Hologic | (p<0.001) | (p<0.001) | (p<0.0001) | DM: 24.2%<br>(p<0.001) | No individual-level data to stratify populations<br>The biopsy rate was higher for DBT+DM group: 1.9% vs. 1.8% (p=0.004) |
| Greenberg (2014)<br>DBT+ DM exams = 20,943<br>DM exams = 38,674<br>Washington, D.C. | Good | Poor (WA HTA, 2014) | Retrospective cohort (2 arm)<br>Community-based multisite radiology practice<br>Mean age: 59.5<br>Test: Selenia Dimensions, Hologic | DBT+ DM: 6.3 (74%)<br>DM: 4.9 (62%)<br>(p=0.035) | DBT+DM: 13.6%<br>DM: 16.2%<br>(p<0.0001) | DBT+DM: 4.6%<br>DM: 3.0%<br>(p=0.0003) | DBT+DM: 22.7%<br>DM: 21.5%<br>(NS) | No follow-up<br>Volunteer bias possible<br>May have overlap with Friedewald (2014)<br>DBT+ DM group had higher biopsy rate (2.6% vs.2.1%, p=0.0003) |
| Haas (2013)<br>DBT+ DM = 6,100 | Good | Poor (WA HTA, 2014) | Retrospective cohort (2 arm)<br>Mean age: 56 | DBT+ DM: 5.7 (69%)<br>DM: 5.2 (68%)<br>(NS) | DBT+ DM: 8.4%<br>DM: 12.0%<br>(p<0.01) | DBT+ DM: 6.8% | NR | No follow-up<br>Women in DBT group had increased risk |

| Author (Year) Study Size Location | Dossier QA | Source QA | Study Design and Population Characteristics | CDR per 1,000 women (% invasive) | Recall Rate (%) | PPV Recall | PPV Biopsy | Comments |
|---|---|---|---|---|---|---|---|---|
| DM = 7, 058<br>Connecticut | | | Test: Selenia Dimensions, Hologic | | | DM: 4.3%[9] | | factors for breast cancer at baseline |
| Houssami (2014)<br>n = 7,292<br>Italy | Good | Good (Melnikow [AHRQ], 2016) | Prospective cohort (1 arm)<br>Population screening program<br>Median age: 58<br>Test: Selenia Dimensions, Hologic | DBT+ DM:7.4<br>DM: 4.8<br>(p<0.001) | DBT+DM: 3.6%<br>DM: 4.2%<br>(NS) | DBT+ DM: 21%<br>DM: 11%[10] | NR | Follow-up 13 months or greater<br>Screen positive if 1 of 2 readers interpreted DM or DBT as abnormal |
| Lourenco (2014)<br>DBT exams = 12,921<br>DM exams = 12,577<br>U.S. | Fair | Poor (WA HTA, 2014) | Retrospective cohort (2 arm), pre<br>Single breast imaging center<br>Mean age: 55.3 DBT, 54.6 DM<br>Test: Selenia Dimensions, Hologic | DBT: 4.6<br>DM: 5.4<br>(NS) | DBT: 6.4%<br>DM: 9.3%<br>(p<0.00001) | DBT: 7.2%<br>DM:5.8%<br>(NS) | DBT: 23.8%<br>DM: 30.2%<br>(sig. NR) | Insufficient follow-up<br>Pre-post design<br>Biopsy rate 1.7% DBT+DM vs 1.6% DM (stat dif NR) |

[9] Center staff calculated by dividing cancers detected by the product of the recall rate and the number of exams, significance not reported
[10] Drawn from AHRQ (2016) report; PPV not reported in original study. Significance not recorded.

| Author (Year) Study Size Location | Dossier QA | Source QA | Study Design and Population Characteristics | CDR per 1,000 women (% invasive) | Recall Rate (%) | PPV Recall | PPV Biopsy | Comments |
|---|---|---|---|---|---|---|---|---|
| McCarthy (2014)<br>DBT+ DM exams = 15,571<br>DM exams = 10, 728<br>Pennsylvania | Fair | Poor (WA HTA, 2014) | Cohort (2 arm)<br>One academic medical center<br>Mean age: 57<br>Test: Selenia Dimensions, Hologic | DBT+ DM: 5.5 (71%)<br>DM: 4.6 (69%)<br>(NS) | DBT+ DM: 8.8%<br>DM: 10.4%<br>(p<0.001) | DBT+ DM: 6.2%<br>DM: 4.4%<br>(p=0.05) | DBT+ DM: 25.7%<br>DM: 24.7%<br>(NS) | Insufficient follow-up<br>Overlap with Friedewald (2014)<br>Pre-post design<br>Biopsy rate for DBT+ DM 2.0% vs. 1.8%, for DM (NS) |
| Rose (2013)<br>DBT+ DM exams = 9,499<br>DM exams = 13,856<br>Texas | Fair | Poor (WA HTA, 2014) | Cohort (2 arm)<br>Multisite, community-based<br>Mean age: NR<br>Test: Selenia Dimensions, Hologic | DBT+ DM: 5.4 (80%)<br>DM: 4.0 (70%)<br>(NS) | DBT+ DM: 5.5%<br>DM: 8.7%<br>(p<0.001) | DBT+ DM: 10.1%<br>DM: 4.7%<br>(p<0.001) | DBT+ DM: 39.8%<br>DM: 26.5%<br>(p=0.06) | No follow-up<br>Pre-post design<br>Biopsy rate 1.1% DBT + DM vs. 1.5% DM (NS) |
| Skaane (2013a)<br>n = 12, 621 exams<br>Norway | Good | Poor (WA HTA, 2014) | Prospective cohort (1 arm)<br>City-wide screening program<br>Mean age: NR | DBT+ DM: 8.0 (80%)<br>DM: 6.1 (73%)<br>(p=0.001) | DBT+ DM: 6.1%<br>DM: 6.7%<br>(p<0.001) | DBT+ DM: 29.1%<br>DM: 28.5%<br>(NS) | NR | Incomplete follow-up<br>Independent double-reading with arbitration prior to recall<br>CDR and recall rate calculated for each |

| Author (Year) Study Size Location | Dossier QA | Source QA | Study Design and Population Characteristics | CDR per 1,000 women (% invasive) | Recall Rate (%) | PPV Recall | PPV Biopsy | Comments |
|---|---|---|---|---|---|---|---|---|
| | | | Test: Selenia Dimensions, Hologic | | | | | image prior to arbitration |
| Skaane (2013b)<br>n = 12, 621 exams<br>Norway | Fair | Poor (WA HTA, 2014) | Prospective cohort (1 arm)<br>City-wide screening program<br>Mean age: 59.3<br>Test: Selenia Dimensions, Hologic | DBT+ DM: 9.4<br>DM: 7.1<br>(p<0.001) | DBT+ DM: 3.7%<br>DM: 2.9%<br>(p<0.001) | DBT+ DM: 24.7%<br>DM: 25.5%<br>(NS) | NR | Incomplete follow-up<br>Independent double-reading with arbitration prior to recall |

Abbreviations: DBT = digital breast tomosynthesis, DM = digital mammography, NS = not significant, NR = not reported, PPV = positive predictive value, stat diff = statistical difference, QA = quality assessment

Harms

*Harm# 1: Over Diagnosis*

Over diagnosis is the detection of cancers that are unlikely to become invasive or cause harm during the course of a person's lifetime. The SRs in this dossier review did not address this harm. Several of the individual cohort studies reported the percentage of cancers that were invasive. The percentage of invasive cancers detected by DBT plus 2D DM ranged from 33% to 80% compared to 50% to 74% for 2D DM alone. Most studies reported comparable rates of invasive cancer detection between modalities, or slightly higher rates of invasive cancer detection for DBT + 2D DM. However, Destounis (2014) reported higher invasive cancer detection using 2D DM alone (50%) compared to combined DBT and 2D DM (33%) (statistical significance not reported). Lang and colleagues (2015) reported slightly lower rates of invasive cancer detected from DBT alone (85%) compared to DM alone (89%).

*Harm #2: False Positive Recalls*

Overall, recall rates are lower when DBT is combined with 2D DM compared to 2D mammography alone and false positive rates are also lower. The five cohort studies reviewed by Nelson and colleagues (2016) reported statistically significantly lower recall rates for DBT and 2D DM compared to mammography alone (Ciatto et al., 2013; Friedewald et al., 2014; Haas et al., 2013; Rose et al., 2013; Skaane et al., 2013a). The WA HTA (2014) TA reported that six studies found statistically significantly lower recall rates for DBT plus 2D DM compared to mammography alone. Table 2 summarizes details for the individual studies included in the SRs and TAs. The overall false positive recall rate in Ciatto (2013) was 5.5%, and a significantly greater number of false positive recalls were from DM readings compared to DBT plus DM (141 vs. 73, $p<0.001$). In the Oslo cohort (Skaane et al., 2013a) the false positive rate before arbitration was 5.3% for DBT plus mammography compared to 6.1% for mammography alone ($p<0.001$). When a pre-arbitration score was based on double readings in the same cohort, the false positive rates were higher (8.5% for DBT plus DM vs. 10.3% for DM alone, $p<0.001$). One prospective cohort detected a higher recall rate for DBT alone compared to 2D mammography alone ($p<0.0001$). Readers participated in an arbitration process prior to recall which may have reduced call-backs (Lang et al., 2015).

Overall, there was a higher probability of DBT plus 2D mammography positive imaging test leading to biopsy-confirmed cancer compared to 2D mammography alone.

*Harm #3: Radiation Dose*

The radiation dose of DBT is similar to that of 2D DM (WA HTA, 2014). When 3D breast imaging is performed in combination with 2D mammography, the radiation dose is doubled. A newer method of DBT imaging technique involves reconstructing 2D images, and therefore contributes no additional radiation exposure beyond the DBT when this method is used. Skaane and

colleagues (2014) compared this technique to DBT plus 2D DM, however, the Center excluded this study because it was comparing two types of DBT. One other study included in the dossier submission assessed DBT with reconstructed 2D images, but was excluded for having a population with a greater than average risk for breast cancer (Zuley et al., 2014).

*Harm #4: Under Diagnosis*

Under diagnosis is defined as failure to detect a disease or condition in a significant proportion of patients. A more sensitive test is less likely to miss cancers, and therefore is less likely to lead to a false negative result. The standard way to assess sensitivity of mammography is to apply a uniform reference standard across modalities and to follow all patients for at least one year to detect interval cancers that may have been missed on screening. Only one study (Houssami et al., 2015), rated as having good methodologic quality by the AHRQ report (Melnikow et al., 2016), applied a reference standard at the end of one year. In this study, the sensitivity of DBT plus 2D DM was 85% (95% CI, 74% to 92%) compared to 54% (95%CI, 42% to 65%) for 2D DM alone. The AHRQ systematic review reported that the sensitivity of 2D DM found in this study is much lower than that of a recent large population-based U.S. study (87%) (AHRQ, 2016).

Additional Harms

Center staff researched the U.S. Food and Drug Administration's Manufacturer and User Facility Device Experience (MAUDE) database on February 8, 2016 and identified two reported harms. One report from 2014 was of itching and breast gland rupture in a patient who underwent DBT with Selenia Dimensions Digital system. The other report from 2013 stated that images were not displaying properly with a Siemens AG Syngo Plaza system.

Ongoing Clinical Trials

Center staff searched Clinicaltrials.gov for any registered trials of DBT for breast cancer screening. See Appendix C for a list of trials that are currently underway or have been recently completed.

Evidence Evaluation – Excluded Studies

Table 3 provides exclusion criteria for submitted articles that were not included in this evaluation.

Table 3. *Submitted References – Reason for Exclusion*

| Citation | Exclusion Criteria |
|---|---|
| Alcusky et al. (2014) | Does not include intervention of interest |
| Bernardi et al. (2014) | Duplicate: Data from STORM Trial (Ciatto et al., 2013) presented with radiologist-specific outcomes |
| Caumo et al. (2015) | Duplicate: Data from STORM trial (Ciatto et al., 2013; Houssami et al., 2014) presented with center-specific outcomes |

| Citation | Exclusion Criteria |
|---|---|
| Gilbert et al. (2015) | Population: Higher risk than asymptomatic screening population |
| Lourenco et al. (2014) | Design: Historically-controlled cohort (included in WA HTA [2014]) |
| Marjolies et al. (2014) | Population: Over 50% either had a history of breast cancer or were at increased risk of breast cancer |
| Rafferty et al. (2014) | Population: Study population enriched with cancer, benign biopsy, and recall cases |
| Rose et al. (2014) | Design: Historically-controlled cohort |
| Sharpe et al. (2015) | Design: Historically-controlled cohort |
| Skaane et al. (2014) | Comparator: Compares two different types of DBT (DBT+ DM and DBT + 2D reconstructed images) |
| Zuley et al. (2014) | Population: Only included women who underwent a biopsy |

Evidence Evaluation – Overall Strength of Body of Evidence by Outcome

Table 4 presents the submitter's assessment of the strength of evidence for the submitted outcomes, as well as the assessment of Center and rationale for this assessment. Evidence that is graded high means that further evidence in unlikely to change our confidence in the estimate of the effect. Moderate strength of evidence means that further evidence is likely to change our confidence in the estimate of the effect and may change the estimate. Low strength of evidence means that further research is very likely to impact our confidence in the estimate of the effect and is likely to change the estimate of effect. Very low strength means that the estimate of the effect is very uncertain (GRADE Working Group, 2004).

The Center for Evidence-based policy uses a Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group approach to strength of evidence to enhance consistency in grading. Randomized controlled trials are initially categorized as having high strength of evidence and observational studies are categorized as having low strength of evidence. The strength rating is downgraded based on limitations including inconsistency of results, some or uncertainty of directness of measurement or population, imprecise or sparse data, and high probability of reporting bias. The grade is increased from low for evidence based on observational studies if there is a strong association,[11]a very strong association,[12] or a dose-response gradient. The grade is also increased is all plausible confounders would have reduced the effect (GRADE Working Group, 2004).

[11] Significant relative risk of >2 or less than <0.5 with no plausible confounders in two or more observational studies

[12] Significant relative risk of >5 or less than <0.2 based on direct evidence with no major threats to validity

Table 4. *Outcomes – Strength of Evidence*

| Outcome | Strength of Evidence Assessment: Submitter | Strength of Evidence Assessment: Center | Rationale |
|---|---|---|---|
| Cancer Detection Rate | High | Low | CDR is consistently higher for DBT + DM compared to DM alone in observational studies. Most studies do not have follow-up to detect interval cancers, and this measure is likely to overestimate the diagnostic accuracy of DBT. (There is a potential for selection bias in several studies that may impact this estimate. There is also is potential temporal confounding in pre-post studies.) |
| Recall Rate | High | Low | Recall rate is consistently lower compared to DBT + DM to DM alone in observational studies. (There is a potential for selection bias due to differences in baseline patient characteristics in several studies that may impact this estimate.) |
| PPV Recall | Moderate | Low | Across observational studies, a positive test is more likely to lead to biopsy-confirmed cancer for DBT + DM compared to DM alone. |
| PPV Biopsy | Moderate | Very low | The measure of PPV of biopsy is imprecise among observational studies. |
| Cost | Moderate -Low | Very low | Cost effectiveness studies to date have based their modeling on individual observational studies and results have been inconsistent. |
| **Harms** | | | |
| Over Diagnosis | Low | Very low | Observational studies report imprecise estimates of percent of invasive cancers detected. Most studies do not have follow-up to detect interval cancers, which may impact the estimate. Heterogeneity in baseline population characteristics is likely to impact this estimate. |
| False Positive Tests (Specificity) | High | Low | Across observational studies, there is a lower recall rate, and those recalled are less likely to be falsely positive. (There is a potential for selection bias due to differences in baseline patient characteristics in several studies that may impact this estimate.) |

| Outcome | Strength of Evidence Assessment | | Rationale |
|---|---|---|---|
| | Submitter | Center | |
| Radiation Dose | Moderate | Low | Observational studies comparing DBT + DM to DM alone have higher radiation doses. This estimate is likely to change as newer methods of performing DBT are utilized. |
| Under Diagnosis (False Negative/ Sensitivity) | High | Very low | Only one observational study had follow-up of interval cancers to assess sensitivity of imaging. This study had a low estimate of DM sensitivity. |

*Section 6: "The service must be cost-effective or cost neutral outside the investigational setting"*
In a poor quality cost-benefit analysis that was included in the dossier submission, Bonafede and colleagues (2015) concluded that a health plan covering one million members, using DBT, could save $2.4 million annually, or $0.20 per member per month through the use of DBT plus 2D DM compared to 2D mammography alone. Authors used a hypothetical population that assumed a lower recall rate for DBT plus 2D mammography (10%) compared to 2D DM alone (15%), and assumed that the DBT plus 2D mammography would detect cancers at an earlier stage. The model also assumed DBT cost an additional $50 per screening test ($4.2 million annually). The savings from fewer follow-up studies after recall was estimated to be $5.5 million in U.S. 2013 dollars, and the cost savings from earlier cancer detection was assumed to be $1.2 million. Limitations of this study include lack of follow-up beyond six months and assumptions that favor DBT. The recall rate for DBT was estimated to be 10% (range, 8% to 12%), and this number was chosen based on the benchmark recall rate advocated by the American College of Radiology and AHRQ. The recall rate for 2D DM was estimated to be over 15% based on claims data. Women who had a diagnostic mammography or breast ultrasound within six months of the initial mammography were assumed to have been recalled. This is likely an overestimate, as women with dense breasts are frequently recalled for ultrasound. In addition, authors assumed an earlier stage of cancer detection based on one study (Skaane et al., 2013b) that was a prospective population-based screening study set in Sweden in which independent double readings were performed. Abnormal results were handled with an arbitration process prior to recall. These methods are unlike U.S. practices in which women are recalled based on a single read. In addition, this study did not follow women for 12 months or longer. Another limitation of this cost-benefit analysis is that a sensitivity analysis was not performed for all inputs and outputs and rates were not discounted.

In a good quality cost-benefit analysis performed in conjunction with the WA HTA (2014) review, DBT plus 2D mammography was determined to cost an additional $56 per patient using a hypothetical average risk cohort based off Washington State census data assuming the DBT

has a slightly better CDR (3.7 vs. 3.6 per 1,000), a smaller number of cancers missed (0.6 vs. 0.7 per 1,000), lower recall rate, and a higher biopsy rate based on the Friedewald (2014) retrospective cohort study. Costs associated with recall included a unilateral diagnostic mammogram in all and an ultrasound for 50% of hypothetical patients, as well as a biopsy in those who were referred. Women presented with interval cancers were assumed to have a diagnostic unilateral mammogram and biopsy. The reduced costs of follow-up imaging were balanced by the increased biopsy costs relative to use of DM alone. The price of DBT was estimated at $57 (based on Centers for Medicare and Medicaid October 2014 ruling) in this study. At a cost of one dollar, DBT would therefore be cost neutral. Results were stable to sensitivity analysis in which sensitivity and specificity values of DBT were increased and CDR was increased relative to DM.

A good quality cost-utility analysis modelled the comparison of biennial DBT plus DM to biennial mammography in a population of asymptomatic women with dense breasts aged 50 to 74 (Lee et al., 2015). In this U.S. population economic model that takes place over a lifetime, there are four interacting processes including breast cancer natural history, detection, treatment, and competing-cause mortality. The population characteristics and mammography performance statistics were drawn from the Breast Cancer Surveillance Consortium. The test characteristics of DBT were drawn from the Oslo cohort (Skaane et al., 2013a), and used in the best-case scenario. This cohort study did not stratify test characteristics by breast density, so overall test measures were used. The additional cost of DBT was estimated to be $50. In the base-case analysis, sensitivity of DBT plus mammography was assumed to be 80% and specificity 92%, which is a moderate improvement over the sensitivity and specificity of mammography (sensitivity 77%; specificity 88%). After 12 screening periods (24 years) and there were 0.5 breast cancer deaths averted, and 405 false readings per 1,000 women averted. The incremental cost per life year gained for combined screening compared to mammography alone was $70,500. The incremental cost per quality adjusted life-year gained was $53,893. This estimate was most sensitive to the cost of DBT, and the sensitivity and specificity of DBT. The incremental cost-effectiveness ratio (ICER) increased to $104,447 when the sensitivity of DBT plus mammography was the same as mammography alone (77%). When specificity of DBT plus mammography was reduced to 90%, the ICER increased to $75,846. The main limitation of this good quality cost utility analysis is that estimates are based on a European cohort that did not have follow-up for interval cancers and arbitration was used prior to recall. The performance characteristics of DBT are likely to be overestimated in this analysis (Lee et al., 2015).

Table 5. *Evidence Review – Economic Studies*

| Study Citation | Dossier QA | Center QA | Study Size (n) | Findings | Limitations / Comments |
|---|---|---|---|---|---|
| Bonafede et al. (2015)<br><br>Study Details<br>Hypothetical population, payer perspective<br><br>Comparison<br><br>DBT+ DM vs. DM alone | Good | Poor | n = 84,549 (hypothetical) | 4,523 women screened with DBT + DM avoided follow-up imaging and biopsy<br><br>$5.5 billion saved in avoiding follow-up costs<br><br>$1.2 million saved from earlier detection of breast cancer<br><br>DBT increased annual cost by $4.2 million<br><br>Costs offset by savings, total savings $2.4 million, cost savings are $0.20 per member per month | Assumptions are likely to over-estimate recall rate of 2D mammography and to overestimate the ability of DBT to detect cancer early compared to DM<br><br>Assumptions drawn from claims data and Skaane (2013b)<br><br>Assumes use of DBT adds $50 to mammography charge<br><br>Funded by Hologic, Inc. |
| WA HTA (2014)<br><br>Study Details<br>Hypothetical population, payer perspective<br><br>Comparison<br>DBT+ DM vs. DM | *Not quality assessed by submitter* | Good | n = 1.3 million (hypothetical) | Savings of reduced recall rates are offset by increased biopsy rates, estimates based on Friedwald et al. (2014)<br><br>Additional cost of DBT per person screening: $56 | Bases assumptions on largest U.S. cohort (Friedewald et al., 2014), which estimates higher biopsy rates DBT+DM compared to DM |
| Lee (2015)<br><br>Study Details<br>Hypothetical population, lifetime model, societal perspective | Good | Good | U.S. population | ICER= $53, 893<br>After 12 screening cycles, 405 per 1,000 false negative screenings averted<br><br>After 12 screening cycle, 0.5 breast-cancer related deaths averted | Based on U.S. population characteristics<br><br>Model most sensitive to DBT cost and sensitivity and specificity of DBT |

| Study Citation | Dossier QA | Center QA | Study Size (n) | Findings | Limitations / Comments |
|---|---|---|---|---|---|
| Comparison DBT+ DM vs. DM | | | | | |

Abbreviations: DBT= digital breast tomosynthesis, DM= digital mammography, NA= not assessed

*Section 7: Other payer coverage of the service*

Center staff searched for policies on the coverage of digital breast tomosynthesis from Aetna, Anthem Blue Cross Blue Shield (BCBS), Cigna, and UnitedHealthCare and the Centers for Medicare and Medicaid Services (CMS). All of the private payers coverage policies reviewed state breast tomosynthesis is investigational and therefore not a covered service (Aetna, Anthem BCBS, Cigna, UnitedHealthCare).

No national or local coverage determinations on digital breast tomosynthesis from CMS were identified. However, in 2015, CMS updated their coverage of screening breast tomosynthesis in response the approval of a new applicable code (current procedural terminology [CPT] 77063). The Centers for Medicare and Medicaid stated that current policies on other mammography apply to breast tomosynthesis. The CPT code 77063 is an add on code, and is only covered in conjunction with a 2D DM (CMS, 2015).

## Summary

Observational cohort studies demonstrate that DBT combined with DM can reduce the recall rate when compared to screening mammography alone. However, this may be partially explained by extra imaging incurred at screening, eliminating the need for additional imaging for some abnormalities noted on mammography. Most studies compare DM to DBT combined with DM, which increases the radiation dose per screening. The biopsy rate is similar to or higher than standard mammography when DBT is added. The CDR is somewhat higher when DBT is combined with mammography, and the frequency of invasive cancers is similar or higher. Most studies do not report results with the use of a comprehensive reference standard and do not report on interval cancers. Therefore, the sensitivity and negative predictive values of the tests are erroneously high. Ongoing studies registered with ClinicalTrails.gov suggest that addition data on interval cancers may be available in the coming years. United States studies that employ a standard approach to breast imaging interpretation and recall, a comprehensive reference standard, reporting on longer-term patient-important outcomes including breast cancer stage, breast cancer recurrence or second breast cancers, and mortality rates are needed.

Private payers do not cover digital breast tomosynthesis. In January of 2015 CMS released a code to cover DBT when performed in conjunction with DM.

## Appendix A. *Search Strategy*

The *MEDLINE®* Search Strategy was adapted from the Agency for Healthcare Research and Quality (2016) systematic review. Studies published after the Hayes (2015) review were included to update the existing systematic reviews.

### *MEDLINE® Search*

Database: Ovid MEDLINE(R) <1946 to January Week 4 2016>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <February 09, 2016>

Search Strategy:

1 Image Processing, Computer-Assisted/ or Radiographic Image Interpretation, Computer-Assisted/ or Tomography, X-Ray Computed/ or Radiographic Image Enhancement/ or Tomography, X-Ray/ or tomosynthesis.mp. or Imaging, Three-Dimensional/

2 exp Breast Neoplasms/

3 (breast adj (neoplasm$ or tumour$ or tumor$ or cancer or carcinoma$ or oncolog$)).mp

4 2 or 3

5 exp Mammography/

6 mammograph$.mp

7 5 or 6

8 exp "Sensitivity and Specificity"/

9 sensitivity.mp

10 specificity.mp

11 ((pre-test or pretest) adj probability).mp

12 ((post-test or posttest) adj probability).mp

13 likelihood ratio.mp

14 8 or 9 or 10 or 11 or 12 or 13

15 1 and 4 and 7 and 14

16 limit 15 to english language

17 limit 16 to (comment or editorial or interview or lectures or letter or news or newspaper article)

18 16 not 17

19 limit 18 to (clinical trial, all or clinical trial, phase i or clinical trial, phase ii or clinical trial, phase iii or clinical trial, phase iv or clinical trial or comparative study or controlled clinical trial or meta analysis or randomized controlled trial or systematic reviews or technical report)

20 limit 19 to yr="2006 -Current"

The search terms, “tomosynthesis,” “3D mammography,” and “three dimensional mammography” were used in the remaining core source searches, which included: Hayes, Inc., the National Institute for Health and Care Excellence (NICE), Cochrane Library, PubMed Health, the Blue Cross/Blue Shield Health Technology Assesesment (HTA) program, the Veterans Administration Technology Assessment Program (VATAP), *BMJ Clinical Evidence*, the Washington State Health Technology Assessment Program, the Agency for Healthcare Research and Quality (AHRQ), and Tufts Cost-Effectiveness Analysis Registry. Systematic reviews that were performed in the last ten years were included. Archived government reports were not included.

## Appendix B. *MEDLINE Results*

Table 1. *MEDLINE Articles Selected for Full Text Review*

| Citation | Included? | Comments/Rationale |
|---|---|---|
| Mercier (2015) | N | Population of women with breast lesions identified on imaging |

## Appendix C. *Ongoing or Recently Completed Clinical Trials of Digital Breast Tomosynthesis*

Randomized Controlled Trial to Evaluate the Efficacy of Digital Breast Tomosynthesis in Reggio Emilia Breast Cancer Screening Program in the 45-74 Age GroupSponsor: Eisai Inc

Sponsor: Azienda Unità Sanitaria Locale Reggio Emilia
Comparators: DBT+ DM vs DM
Design: Parallel randomized open label
Primary Outcomes: cumulative incidence of T2+ cancers after screening, incidence of interval cancers
Primary Completion: December 2016

Assessment of Digital Breast Tomosynthesis (DBT) in the Screening Environment: a Prospective Study

Sponsor: University of Pittsburg
Comparators: DBT+ DM vs. DM
Design: Single arm prospective cohort
Primary Outcomes: recall rate
Primary Completion: May 2014

A Study to Determine Patient Benefit of Tomosynthesis in Screening Mammography

Sponsor: Hologic, Inc.
Comparators: DBT+ DM vs. DM
Design: Parallel randomized open label
Primary Outcomes: Interpretation time of scan
Primary Completion: April 2014
Status: Completed

Tomosynthesis Mammography Imaging Screening Trial Lead-in

Sponsor: Sunnybrook Health Sciences Centre
Comparators: DBT+ DM vs. DM
Design: Parallel randomized open label
Primary Outcomes: Diagnostic accuracy using are under the curve score generated by receiver operator characteristic analysis
Primary Completion: November 2018

A Multicenter, Controlled Clinical Trial to Evaluate the Hologic Tomosynthesis Mammography

Sponsor: Hologic, Inc.
Comparators: DBT+ DM vs. DM
Design: Parallel non-randomized open label

Primary Outcomes: Diagnostic accuracy using are under the curve score generated by receiver operator characteristic analysis
Primary Completion: December 2012
Status: Active, not recruiting

Digital Breast Tomosynthesis in the Oslo Mammography Screening Program

Sponsor: Oslo University Hospital
Comparators: DBT+ DM vs. DM
Design: Prospective cohort
Primary Outcomes: Screening performance indicators
Primary Completion: December 2012
Status: Active, not recruiting

Digital Breast Tomosynthesis vs. Digital Mammography: A National Multicenter Trial

Sponsor: Medical University of Vienna
Comparators: DBT vs. DM
Design: Prospective cohort
Primary Outcomes: Specificity
Primary Completion: September 2012
Status: Recruiting (not verified recently)

Assessment of Diagnostic Accuracy and Performance of Digital Breast Tomosynthesis Compared to Mammography (ADAPT Trial) ADAPT-SCR: Recruitment Plan for Asymptomatic Women Undergoing Screening Mammography

Sponsor: GE Healthcare
Comparators: DBT vs. DM
Design: non-randomized crossover open label
Primary Outcome: cancer status
Primary Completion: November 2017

Assessment of Diagnostic Accuracy and Performance of Digital Breast Tomosynthesis Compared to Mammography (ADAPT Trial) ADAPT-BX: Recruitment Plan for Initially Asymptomatic Women Referred for Breast Biopsy

Sponsor: GE Healthcare
Comparators: DBT vs. DM
Design: non-randomized, cross-over, open label
Primary outcome: cancer status
Primary completion: December 2016

Comparison of Full-Field Digital Mammography With Digital Breast Tomosynthesis Image Acquisition in Relation to Screening Call-Back Rate

Sponsor: American College of Radiology Imaging Network
Comparators: DBT plus low dose mediolateral oblique mammography view compared to mammography
Design: non-randomized, parallel, open label
Primary outcome: recall rates
Primary completion: June 2012
Status: Active, not recruiting

Malmö Breast Tomosynthesis Screening Trial

Sponsor: Region Skane
Comparators: DBT plus mediolateral oblique mammography view compared to mammography
Design: single group, open-label
Primary outcome: number of breast cancers detected
Primary completion: December, 2017

Digital Breast Tomosynthesis Versus Digital Mammography in a Population-based Screening Program. A Controlled Randomized Multicenter Trial

Sponsor: Centro di Riferimento per l'Epidemiologia e la Prev. Oncologica Piemonte
Comparators: DBT vs. DM
Design: randomized, parallel, single blind (subject)
Primary outcomes: rates of cancers after first screening round. Data or interval cancers (24 months after initial screen) and advanced screen-detected cancers at the subsequent screen among participants will be collected.
Primary completion: December 2017

A Comparison of Recall Rates Between Conventional 2d Mammography and 2d Plus 3d (Tomosynthesis) Mammography in a Screening Population

Sponsor: Rose Imaging Specialists, P.A.
Comparators: DBT+ DM vs. DM
Design: retrospective cohort
Primary outcome: recall rate
Primary completion: not reported

A Multicenter Study to Test Digital Breast Tomosynthesis (DBT) Compared to Full-Field Digital Mammography (FFDM) in Detecting Breast Cancer. Part 1. Women Undergoing Screening Mammography

Sponsor: GE Healthcare

Comparators: DBT vs. DM
Design: prospective cohort
Primary outcome: diagnostic performance
Primary completion: June 2009
Status: completed

A Multi-Reader Multi-Case Controlled Clinical Trial to Assess the Adequacy of the Fujifilm Full Field Digital Mammography (FFDM) and Digital Breast Tomosynthesis (DBT) Reader Training Program - A Pilot Study

Sponsor: Fujifilm Medical Systems USA, Inc.
Comparators: DBT+ DM vs. DM
Design: randomized, crossover, open-label
Primary outcome: cancer detection rate
Primary completion: August 2015
Status: completed

## Appendix D. *Quality Assessment Forms*

Table 1a. *Systematic Reviews Quality Assessment*

| Risk of Bias Assessment Criteria | Melnikow (2016) | | Nelson (2015) | | WA HTA (2014) | |
|---|---|---|---|---|---|---|
| | Submitter | Center | Submitter | Center | Submitter | Center |
| 1.1 The study addresses an appropriate and clearly focused question. | *Not included in dossier submission* | Yes | *Not included in dossier submission* | Yes | Yes | Yes |
| 1.2 An adequate description of the methodology used is included, and the methods used are appropriate to the question. | | Yes | | Yes | Yes | Yes |
| 1.3 The literature search is sufficiently rigorous to identify all the relevant studies. | | Yes | | Yes | Yes | Yes |
| 1.4 The criteria used to select articles for inclusion is appropriate. | | Yes | | Yes | Yes | Yes<br>However, they include studies that have a follow-up of less than one year. This may miss interval cancers |
| 1.5 Study quality is assessed and taken into account. | | Yes | | Yes | Yes | Yes |
| 1.6 There are enough similarities between the studies selected to make combining them reasonable. | | Yes<br>One study | | Yes | Yes | Yes |
| 1.7 There is a conflict of interest statement. | | Yes | | Yes | Unclear | No |
| 1.8 There is a description of the source(s) of funding. | | Yes | | Yes | Unclear | Yes |
| **2.1 How well was the study done to minimize bias?** | | **Good** | | **Good** | **Good** | **Good** |
| 2.2 Are the results of this study directly applicable to the patient group targeted by this key question? | | Yes | | Yes | Yes | Yes |
| 2.3 Comments | | --- | | --- | --- | --- |

Table 2. *Diagnostic Test Accuracy Study Quality Appraisal*

| Risk of Bias Assessment Criteria | Durand (2014) | | Lang (2015) | |
|---|---|---|---|---|
| | Submitter | Center | Submitter | Center |
| 1.1 The spectrum of patients is representative of the patients who will receive the test in practice. | *Submitter used the Cohort Quality Appraisal Checklist, and rated the study as Good.* | Yes | Yes | Yes<br>Asymptomatic, urban Swedish population |
| 1.2 Selection criteria are clearly described. | | Yes | Yes | Yes |
| 1.3 The reference standard is likely to classify the condition correctly. | | Yes | Yes | Yes<br>Biopsy |
| 1.4 The period between reference standard and index test is short enough to be reasonably sure that the target condition did not change between the two tests. | | Yes | Yes | Yes |
| 1.5 The whole sample, or a random selection of the sample, received verification using a reference standard of diagnosis. | | No | Yes | No |
| 1.6 Patients received the same reference standard regardless of the index test result. | | No | Yes | No<br>No reference standard applied to normal tests (no follow-up) |
| 1.7 The reference standard was independent of the index test (i.e. the index test did not form part of the reference standard). | | Yes | Yes | Yes |
| 1.8 The execution of the index test was described in sufficient detail to permit replication of the test. | | Yes | Yes | Yes |
| 1.9 The execution of the reference standard was described in sufficient detail to permit replication of the test. | | Yes | Yes | Yes |
| 1.10 Index test results were interpreted without knowledge of the results of the reference standard. | | Yes | Yes | Yes |
| 1.11 Reference standard results were interpreted without knowledge of the results of the index test. | | No | Yes | No |
| 1.12 Uninterpretable or intermediate test results are reported. | | n/a | No | n/a |
| 1.13 An explanation is provided for withdrawals from the study. | | n/a<br>No follow-up | No | n/a<br>No follow-up |
| 1.14 Competing interests of members have been recorded and addressed. | | Yes<br>Several authors have received consulting money from Hologic, Inc., one | Yes | Yes<br>Authors have relationship with Siemans (manufacturer of tomosynthesis test equipment) |

| Risk of Bias Assessment Criteria | Durand (2014) | | Lang (2015) | |
|---|---|---|---|---|
| | Submitter | Center | Submitter | Center |
| | | author has an affiliation with Siemans | | |
| 1.15 Views of funding body have not influenced the content of the study. | | Unclear<br>Not reported | Yes | No<br>Funded by Siemans |
| **2.1 How well was the study done to minimize bias?** | | **Poor** | **Good** | **Poor** |
| 2.2 If coded as Fair or Poor what is the likely direction in which bias might affect the study results? | | There are baseline differences in the study population that is likely to bias results. Also, there is no follow-up and reference standard is not applied to negative tests. | n/a | There is a lack of uniform reference standard applied to all test results. Interval breast cancers cannot be identified, and therefore sensitivity and negative predictive value cannot be detected. Cancer detection rates are likely overestimated. |
| 2.3 Are the results of this study directly applicable to the patient group targeted by this topic? | | Yes | Yes | Yes |
| 2.4 Other reviewer comments: | | --- | Post-arbitration recall rate for the DM population was biased because BT information was used during arbitration. Thus, the recall rate for the DM population would have likely been higher if BT information were not available. | Arbitration process likely resulted in lower recall rate than seen in U.S., where practice is to recall after one read |

Table 3. *Economic Study Quality Appraisal*

| Risk of Bias Assessment Criteria | Bonafede (2015) | | Lee (2015) | | WA HTA (2014) | |
|---|---|---|---|---|---|---|
| | Submitter | Center | Submitter | Center | Submitter | Center |
| 1.1 The results of this study are directly applicable to the patient group targeted by this key question. | Yes | Yes | Yes | Yes | *See Table 1. Systematic Review Quality Assessment for overall quality assessment by submitter. Individual quality assessment not included by submitter.* | Yes<br>However, WA state population data used in creating model, likely to differ with NY demographics |
| 1.2 The healthcare system in which the study was conducted is sufficiently similar to the system of interest in the topic key question(s). | Yes | Yes<br>U.S. Healthcare system | Yes | Yes | | Yes |
| 2.1 The research question is well described. | Yes | Yes | Yes | Yes | | Yes |
| 2.2 The economic importance of the research question is stated. | Yes | Yes | Yes | Yes | | Yes |
| 2.3 The perspective(s) of the analysis are clearly stated and justified (e.g. healthcare system, society, provider institution, professional organization, patient group). | Yes | Yes<br>Healthcare payer perspective | Yes | Yes<br>Society | | Yes<br>State payer |
| 2.4 The form of economic evaluation is stated and justified in relation to the questions addressed. | Yes | Yes | No | Yes | | Yes |
| 2.5 Details of the methods of synthesis or meta-analysis of estimates are given (if based on a synthesis of a number of effectiveness studies). *or*<br>Details of the design and results of effectiveness study are given (if based on a single study). | n/a | No<br>Details partially given. Methods of underlying study no sufficiently described, needed to reference article | No | Yes | | Yes |
| 2.6 Estimates of effectiveness are used appropriately. | Yes | Yes | Yes | Yes | | Yes<br>Based on Friedewald (2014) – largest U.S. cohort study |
| 2.7 Methods to value health states and other benefits are stated. | Yes | n/a | n/a | Yes | | Yes |
| 2.8 Outcomes are used appropriately. | Yes | Yes | Yes | Yes | | Yes |
| 2.9 The primary outcome measure for the economic evaluation is clearly stated. | Yes | Yes | No | Yes | | Yes |

| Risk of Bias Assessment Criteria | Bonafede (2015) | | Lee (2015) | | WA HTA (2014) | |
|---|---|---|---|---|---|---|
| | Submitter | Center | Submitter | Center | Submitter | Center |
| 2.10 Details of the subjects from whom valuations were obtained are given. | Yes | Yes | Yes | Yes | | Yes |
| 2.11 Competing alternatives are clearly described. | Yes | Yes | No | Yes | | Yes |
| 2.12 All important and relevant costs for each alternative are identified. | Yes | Yes | No | Yes | | Yes |
| 2.13 Methods for the estimation of quantities and unit costs are described. | Yes | Yes | Yes | Yes | | Yes |
| 2.14 Quantities of resource use are reported separately from their unit costs. | Yes | Yes | Yes | No | | Yes |
| 2.15 Productivity changes (if included) are reported separately. | n/a | No | n/a | No | | No |
| 2.16 The choice of model used and the key parameters on which it is based are justified. | Yes | No<br>Assumed recall rate of 10% for DBT + DM as this corresponds with AHRQ benchmark. Recalls for DM drawn from claims database (DM or breast ultrasound within 6 months of initial mammography ≥ 15.35%. Assumed cancers detected at an earlier stage based on a Swedish study (Skaane, 2013b) in which a double reading with arbitration process was used to compare 2DM to 2DM + DBT. This process is not comparable to the practice in the U.S. This study also had incomplete follow-up. | No | Yes | | Yes |
| 2.17 All costs are measured appropriately in physical units. | Yes | Yes | Yes | Yes | | Yes |
| 2.18 Costs are valued appropriately. | Yes | Unclear | Yes | Yes | | Yes |

| Risk of Bias Assessment Criteria | Bonafede (2015) | | Lee (2015) | | WA HTA (2014) | |
|---|---|---|---|---|---|---|
| | Submitter | Center | Submitter | Center | Submitter | Center |
| | | Assumed DBT + 2DM costs $50 more than 2DM | | | | |
| 2.19 Outcomes are valued appropriately. | Yes | Yes | Yes | Yes | | Yes |
| 2.20 The time horizon is sufficiently long enough to reflect all important differences in costs and outcomes. | Yes | No<br>Follow-up is for 6 months, interval cancers not assessed | Yes | Yes | | Yes<br>One year |
| 2.21 The discount rate(s) is stated. | Yes | No | n/a | Yes | | No |
| 2.22 An explanation is given if costs and benefits are not discounted. | n/a | No | n/a | n/a | | No |
| 2.23 The choice of discount rate(s) is justified. | No | n/a | n/a | Yes | | n/a |
| 2.24 All future costs and outcomes are discounted appropriately. | Yes | No | n/a | Yes | | No |
| 2.25 Details of currency of price adjustments for inflation or currency conversion are given. | Yes | Yes<br>2013 U.S. dollars | n/a | Yes | | n/a<br>2014 U.S. dollars |
| 2.26 Incremental analysis is reported or it can be calculated from the data. | Yes | Yes | n/a | Yes | | Yes |
| 2.27 Details of the statistical tests and confidence intervals are given for stochastic data. | No | No | n/a | No | | Yes |
| 2.28 Major outcomes are presented in a disaggregated as well as aggregated form. | Yes | Yes | Yes | Yes | | Yes |
| 2.29 Conclusions follow from the data reported. | Yes | Yes | Yes | Yes | | Yes |
| 2.30 Conclusions are accompanied by the appropriate caveats. | Yes | No | Yes | Yes | | Yes |
| 3.1 The approach to sensitivity analysis is given. | Yes | No | Yes | Yes | | Yes |
| 3.2 All important and relevant costs for each alternative are identified. | Yes | No | Yes | Yes | | Yes |
| 3.3 An incremental analysis of costs and outcomes of alternatives is performed. | Yes | No | n/a | Yes | | Yes |
| 3.4 The choice of variables for sensitivity analysis is justified. | No | No | Yes | Yes | | Yes |
| 3.5 All important variables, whose values are uncertain, are appropriately subjected to sensitivity analysis. | Yes | No | No | Yes | | Yes |
| 3.6 The ranges over which the variables are varied are justified. | No | No | No | Yes | | Yes |

| Risk of Bias Assessment Criteria | Bonafede (2015) | | Lee (2015) | | WA HTA (2014) | |
|---|---|---|---|---|---|---|
| | Submitter | Center | Submitter | Center | Submitter | Center |
| 4.1 Competing interests of members have been recorded and addressed. | Yes | Yes | Yes | Yes | | Yes |
| 4.2 Views of funding body have not influenced the content of the study. | Yes | No<br>Funded by Hologic, Inc | Yes | Unclear<br>Partially sponsored by GE, some authors affiliated with GE | | Yes |
| **5.1 How well was the study done to minimize bias?** | **Good** | **Poor** | **Good** | **Good** | | **Good** |
| 5.2 If coded as fair or poor, what is the likely direction in which bias might affect the study results? | n/a | Likely to be biased in favor of DBT. Assumptions are based on claims data which may overestimate the recall rate (many women may have ultrasound for dense breasts), and one study that is likely to overestimate the cancer detection rate. It used double reading with arbitration. | n/a | n/a | | n/a |
| 5.3 Other reviewer comments: | --- | --- | --- | --- | | --- |

## References


Bonafede, M. M., Kalra, V. B., Miller, J. D., & Fajardo, L. L. (2015). Value analysis of digital breast tomosynthesis for breast cancer screening in a commercially-insured US population. *ClinicoEconomics and Outcomes Research, 7*, 53-63. doi: 10.2147/CEOR.S76167

Ciatto, S., Houssami, N., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S., … Macaskill, P. (2013). Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): A prospective comparision study. *Lancet Oncology, 14*(7), 583-589. http://dx.doi.org/10.1016/S1470-2045(13)70134-7

Destounis, S., Arieno, A., & Morgan, R. (2014). Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. *Journal of Clinical Imaging Science, 4*(1), 1-6. DOI: 10.4103/2156-7514.127838

Durand, M. A., Hass, B. M., Yao, X., Geisel, J. L., Raghu, M., Hooley, R. J., … Philpotts, L. E. (2014). Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology, 274*(1), 85-92.

Friedewald, S. M., Rafferty, E. A., Rose, S. L., Durand, M. A., Plecha, D. M., Greenberg, J. S., … Conant, E. F. (2014). Breast cancer screening using tomosynthesis in combination with digital mammography. *Journal of the American Medical Association, 311*(24), 2499-2507. doi:10.1001/jama.2014.6095

GRADE Working Group. (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal, 328*, 1-8.

Greenberg, J. S., Javitt, M. C., Katzen, J., Michael, S., & Holland, A. E. (2014). Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *Woman's Imaging, 203*(3), 1-7. DOI:10.2214/AJR.14.12642

Haas, B. M., Kalra, V., Geisel, J., Raghu, M., Durand, M., & Philpotts, L. E. (2013). Comparison of tomosynthesis plus digital mammography and digital mammography along for breast cancer screening. *Radiology, 269*(3), 694-700.

Houssami, N., Macaskill, P., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S., … Ciatto, S. (2014). Breast cancer screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading – Evidence to guide future screening strategies. *European Journal of Cancer, 50*(10), 1799-807.

Lang, K., Andersson, I., Rosso, A., Tingberg, A., Timberg, P., & Sackrisson, S. (2015). Performance of one-view breast tomosynthesis as a stand-alone breast cancer screening modality: Results from the Malmo Breast Tomosynthesis Screening Trial, a population-based study. *European Radiology, 26*(1), 184-90 . DOI 10.1007/s00330-015-3803-3

Lee, C. I., Cevik, M., Alagoz, O., Sprague, B. L., Tosteson, A. N. A., Miglioretti, D. L., … Lehman, C. D. (2015). Comparative effectiveness of combined digital mammography and tomosynthesis screening for women with dense breasts. *Radiology, 274*(3), 772-80.

Lourenco, A. P., Barry-Brooks, M., Baird, G., Tuttle, A., & Mainiero, M. B. (2014). Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology,* 274(2), 337-42. DOI: http://dx.doi.org/10.1148/radiol.14140317

McCarthy, A. M., Kontos, D., Synnestvedt, M., Tan, K. S., Heitjan, D. F., Schnall, M., & Conant, E. F. (2014). Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *Journal of the National Cancer Institute, 106*(11), 1-7. DOI:10.1093/jnci/dju316

Melnikow, J., Fentron, J. J., Miglioretti, D., Whitlock, E. P., & Weyrich, M. S. (2016). *Screening for breast cancer with digital breast tomosynthesis.* Rockville, MD: Agency for Healthcare Research and Quality.

Nelson, H. D., Pappas, M., Cantor, A., Griffin, H., Daeges, M., & Humphrey, L. (2016). Harms of breast cancer screening: Systematic review to update the 2009 U.S. Preventive Services Task Force recommendation. *Annals of Internal Medicine, 164*(4), 256-267 . doi:10.7326/M15-0970

New York Department of Health. (2015). Dossier methods guidance. Retrieved from http://www.health.ny.gov/health_care/medicaid/redesign/docs/dossier_methods_guidance.pdf

Rose, S. L., Tidwell, A. L., Bujnoch, L. J., Kushwaha, A. C., Nordmann, A. S., & Sexton, R. Jr. (2013). Implementation of breast tomosynthesis in a routine screening practice: An observational study. *Women's Imaging, 200*(6), 1401-1408. DOI:10.2214/AJR.12.9672

Skaane, P., Bandos, A. I., Gullien, R., Eben, E. B., Ekseth, U., Haakenaasen, U., … Gur, D. (2013a). Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology, 267*(1), 47-56.

Skaane, P., Bandos, A. I., Gullien, R., Eben, E. B., Ekseth, U., Haakenaasen, U., … Hofvind, S. (2013b). 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. *European Radiology, 23*(6), 2061-2071. DOI 10.1007/s00330-013-2820-3

Skaane, P., Bandos, A. I., Eben, E. B., Jebsen, I. N., Krager, M., Haakenaasen, U., … Gullien, R. (2014). Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: Comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology, 271*(3), 655-663.

Washington Health Technology Assessment Program (WA HTA). (2014). *Appropriate imaging for breast cancer screening in special populations.* Olympia, WA: WA HTA. Retrieved from http://www.hca.wa.gov/hta/documents/breast_imaging_draft_rpt_comments_response_121014.pdf

## About the Center for Evidence-based Policy

The Center for Evidence-based Policy (Center) is recognized as a national leader in evidence-based decision making and policy design. The Center understands the needs of policymakers and supports public organizations by providing reliable information to guide decisions, maximize existing resources, improve health outcomes, and reduce unnecessary costs. The Center specializes in ensuring diverse and relevant perspectives are considered, and appropriate resources are leveraged to strategically address complex policy issues with high-quality evidence and collaboration. The Center is based at Oregon Health & Science University in Portland, Oregon. Further information about the Center is available at www.ohsu.edu/policycenter.

**Suggested citation: Pensa, M., Thielke, A., & King, V. (2016). *Breast tomosynthesis – dossier review.* Portland, OR: Center for Evidence-based Policy, Oregon Health & Science University.**

Conflict of Interest Disclosures: No authors have conflicts of interest to disclose. All authors have completed and submitted the Oregon Health & Science University form for Disclosure of Potential Conflicts of Interest, and none were reported.

Funding/Support: This research was funded by the Center for Evidence-based Policy's Medicaid Evidence-based Decisions project at Oregon Health & Science University.