CADTH

CADTH RAPID RESPONSE REPORT:
SUMMARY WITH CRITICAL APPRAISAL

Digital Breast Tomosynthesis for the Screening and Diagnosis of Breast Cancer: A Review of the Diagnostic Accuracy, Cost-Effectiveness and Guidelines

Service Line:	Rapid Response Service
Version:	1.0
Publication Date:	October 28, 2019
Report Length:	41 Pages

CADTH

Authors: Wade Thompson, Charlene Argaez

Cite As: Digital Breast Tomosynthesis for the Screening and Diagnosis of Breast Cancer: A Review of the Diagnostic Accuracy, Cost-Effectiveness and Guidelines. Ottawa: CADTH; 2019 Oct. (CADTH rapid response report: summary with critical appraisal).

ISSN: 1922-8147 (online)

About CADTH: CADTH is an independent, not-for-profit organization responsible for providing Canada's health care decision-makers with objective evidence to help make informed decisions about the optimal use of drugs, medical devices, diagnostics, and procedures in our health care system.

Funding: CADTH receives funding from Canada's federal, provincial, and territorial governments, with the exception of Quebec.

Questions or requests for information about this report can be directed to Requests@CADTH.ca

Abbreviations

2D	Two dimensional
3D	Three dimensional
BIRADS	Breast imaging reporting and data system
CDR	Cancer detection rate
CI	Confidence interval
DBT	Digital breast tomosynthesis
DM	Digital mammography
DOR	Diagnostic odds ratio
NLR	Negative likelihood ratio
PLR	Positive likelihood ratio
PPV	Positive predictive value
QALY	Quality adjusted life-year
RCT	Randomized controlled trial
RR	Relative risk
SR	Systematic review

Context and Policy Issues

Among Canadian women, breast cancer is the most common type of cancer and the second leading cause of death from cancer.[1] As of 2019, it is estimated that approximately 1 in 8 Canadian women develop breast cancer during their lifetime, while around 1 in 33 Canadian women die from breast cancer.[1] Breast cancer most commonly affects women over the age of 40,[2] with the median age at diagnosis 62 years (based on data from the United States).[3]

Screening for breast cancer in women aims for early diagnosis of the disease, the possibility of more effective and less invasive treatment, and ultimately improved outcomes.[4] Some women may display symptoms of breast cancer or have suspicious signs when undergoing screening, which may necessitate diagnostic testing.[5,6] An effective screening or diagnostic technique should demonstrate a positive effect on clinical outcomes (such as mortality) and should not lead to overdiagnosis (e.g. diagnosis and subsequent treatment of cancer would not have caused harm if not treated), or produce high numbers of false-positive results.[7]

Canada has had breast cancer screening programs in place since the 1990s.[4] Various techniques may be used for screening and diagnosis of breast cancer. The current standard for screening is digital mammography (DM; also called 2D mammography), which uses X-rays and digital detection.[5] This technique is also used for diagnosis in women with symptoms or suspicious results after a screening mammogram.[6] An emerging technology is digital breast tomosynthesis (DBT; also called 3D digital tomosynthesis).[8] This technique also uses X-ray and digital images; however, it involves taking multiple images from multiple angles to produce 3D images of the breast.[8] It is also possible to obtain synthesized 2D images using DBT.[8] DBT can be used alone or in combination with DM. Screening using DM alone has led to reductions in breast cancer mortality, though there are still concerns related to overdiagnosis and false positive results, as well as high recall rates (the number of women needing follow-up testing after an initial screen).[4,9,10] The goal of DBT alone or in combination with DM is therefore to improve accuracy of screening and diagnosis, and ultimately lead to improved outcomes for patients, such as reductions in mortality rate and overdiagnosis.[7]

The aim of the report is to summarize the clinical utility, diagnostic accuracy, and cost-effectiveness of DBT with or without DM compared to DM alone, and to summarize existing guidelines on this topic. This report is based on a previous CADTH Summary of Abstracts titled “Digital Tomosynthesis for the Screening and Diagnosis of Breast Cancer: Diagnostic Accuracy, Cost-Effectiveness, and Guidelines”.[11]

Research Questions

1. What is the clinical utility of 3D digital tomosynthesis with or without 2D mammography compared with 2D mammography alone for breast cancer screening or diagnosis?
2. What is the diagnostic accuracy of 3D digital tomosynthesis with or without 2D mammography compared with 2D mammography alone for breast cancer screening or diagnosis?
3. What is the cost effectiveness of 3D digital tomosynthesis with or without 2D mammography compared with 2D mammography alone for breast cancer screening or diagnosis?
4. What are the evidence-based guidelines regarding the use of 3D digital tomosynthesis for breast cancer screening and diagnosis?

Key Findings

Based on seven systematic reviews of nonrandomized studies, digital breast tomosynthesis in combination with digital mammography may improve detection rate and recall rate of breast cancer compared to digital mammography alone for screening, though there was heterogeneity in evidence with respect to the type of cancer (invasive versus noninvasive) and screening setting (frequency of screening, number of reads). Results were conflicting in the two eligible randomized controlled trials. Both found no benefit on detection rate for digital breast tomosynthesis in addition to digital mammography for screening, while one randomized controlled trial found benefit for recall rate and the other did not.

There was limited evidence in the diagnostic setting. For diagnosis, two systematic reviews provided narrative comparisons of digital breast tomosynthesis alone or in combination with digital mammography to digital mammography alone. Both reported that digital breast tomosynthesis improved sensitivity, though results were conflicting for specificity. One systematic review found that digital breast tomosynthesis alone or in combination with digital mammography improved the detection rate of breast cancer for women with dense breasts in a diagnostic setting.

No evidence on the clinical effectiveness or harms of digital breast tomosynthesis was identified. As such, the benefits and harms of digital breast tomosynthesis for screening and diagnosis are unclear.

One cost-effectiveness study conducted in the United States found that digital breast tomosynthesis in combination with digital mammography was cost-effective compared to digital mammography alone in women age 40 to 79. Digital breast tomosynthesis and digital mammography was most cost-effective in those age 40 to 49, compared to those age 50 to 59, 60 to 69, and 70 or older.

Two guidelines were identified, both of which recommended against using digital breast tomosynthesis for screening of breast cancer in asymptomatic women not at high risk of

breast cancer. Recommendations were based on the insufficient evidence for digital breast tomosynthesis on benefits and harms.

Methods

Literature Search Methods

This report makes use of a literature search developed for a previous CADTH report.[11] The original literature search was conducted in May 2019 on key resources including Medline via OVID, The Cochrane Library, University of York Centre for Reviews and Dissemination (CRD) databases, Canadian and major international health technology agencies, as well as a focused Internet search. The search strategy was comprised of both controlled vocabulary, such as the National Library of Medicine's MeSH (Medical Subject Headings), and keywords. The main search concepts were digital tomosynthesis and breast cancer. Search filters were applied to limit retrieval to health technology assessments, systematic reviews, meta-analyses, or network meta-analyses, randomized controlled trials or controlled clinical trials, economic studies or guidelines. An additional focused search with no search filters was also conducted. Where possible, retrieval was limited to the human population. The initial searches were also limited to English-language documents published between January 1, 2014 and May 23, 2019.

Selection Criteria and Methods

One reviewer screened citations and selected studies for full-text review, while another reviewer reviewed the full-text versions for inclusion. In the first level of screening, titles and abstracts were reviewed and potentially relevant articles were retrieved and assessed for inclusion. The final selection of full-text articles was based on the inclusion criteria presented in Table 1.

Table 1: Selection Criteria

Population	Adult women, subgroups: • Adult women ages 50 to 74 • Adult women with low breast density • Adult women with high breast density
Intervention	3D digital tomosynthesis with or without 2D mammography
Comparator	2D mammography alone
Outcomes	Question 1: Clinical utility: Safety, adverse events (e.g. radiation) Question 2: Diagnostic accuracy (e.g., accuracy, sensitivity, specificity, detection rates) Question 3: Cost effectiveness Question 4: Evidence-based guidelines
Study Designs	Health technology assessments, systematic reviews/meta-analyses, randomized controlled trials, economic evaluations, guidelines

Exclusion Criteria

Articles were excluded if they did not meet the selection criteria outlined in Table 1 or they were duplicate publications. Guidelines with unclear methodology were also excluded. The Summary of Abstracts on which this report is based included non-randomized studies; however, for the purposes of this report non-randomized studies were excluded. Economic

studies that were basic costing studies that did include a cost-effectiveness or cost-utility analysis were also excluded.

Critical Appraisal of Individual Studies

The included systematic reviews were critically appraised by one reviewer using AMSTAR 2,[12] RCTs that were diagnostic studies were appraised using QUADAS-2,[13] the economic study was appraised using the Drummond Checklist,[14] and guidelines were evaluated using AGREE II.[15] A review of strengths and limitations was provided narratively.

Summary of Evidence

Quantity of Research Available

A total of 581 citations were identified in the literature search. Following screening of titles and abstracts, 14 were retrieved for full-text review. Of these potentially relevant articles, two publications were excluded, and 12 publications met the inclusion criteria and were included in this report. These comprised seven systematic reviews (SR), two randomized controlled trials (RCTs), one economic evaluation, and two evidence-based guidelines. Appendix 1 presents the PRISMA flowchart of the study selection.

Summary of Study Characteristics

Study Design

A total of seven SRs were identified,[16-22] published between 2014 and 2018. The dates the searches were conducted ranged from February 2013 to July 2017. The primary studies in the eligible SRs were nonrandomized prospective and retrospective studies. There was overlap in eligible primary studies among the systematic reviews (see Appendix 5).

Two randomized controlled trials (RCTs) were identified.[10,23] One employed a parallel group design[10] and the other used a cross-over design[23] where groups received DBT and DM at baseline, then DM alone one year later, or vice versa.

One cost-effectiveness study[24] was identified, which evaluated DBT plus DM versus DM alone for screening of breast cancer in women age 40 to 79. This study was conducted from the perspective of a federal payer (in the United States) and used a lifetime time horizon. Inputs came from Medicare reimbursement data and an American screening study.[25] A decision tree model was used for analysis. The authors assumed there would be no false negatives with DBT since it was treated as a reference standard. They also assumed no patients would be lost to follow-up and that any recall testing was only radiologic (not biopsy).

Two guidelines were eligible, both of which were breast cancer screening guidelines published in 2018.[4,26] One was a Canadian guideline from the Canadian Task Force on Preventive Health Care.[4] These authors conducted a systematic review of systematic reviews to identify relevant evidence, and the GRADE approach to develop recommendations. Recommendations were based on voting. A Brazilian guideline was also identified.[26] The body responsible for producing the guideline is not clear from the guideline text. Recommendations were based on SRs of evidence. Details surrounding evidence synthesis, development of recommendations, and voting on recommendations, were unclear as the accompanying methods document was only available in Portuguese.

Further details regarding eligible studies are in Appendix 2.

Country of Origin

Systematic reviews were conducted by groups in Australia,[16,22] South Korea,[17] United Kingdom,[19] China,[21] Netherlands,[18] and Spain.[20] RCTs were conducted in Norway[10] and United Kingdom.[23] Guidelines were from Canada[4] and Brazil.[26]

Patient Population

Four of the eligible SRs examined DBT for screening only.[16,17,19,22] Women in the studies from these SRs were those attending breast cancer screening programs. They were asymptomatic. The mean/median age of women in studies from the screening SRs ranged from 56 years to 59 years. One SR specifically evaluated DBT for diagnosis.[21] The characteristics of the women in this SR were not described by the authors. Two of the SRs included both screening and diagnostic studies.[18,20] One of these SRs included only women with dense breasts (Breast Imaging Reporting and Data System [BIRADS] 3 or 4/c or d) who were asymptomatic.[18] The mean/median age of women in this SR ranged from 49 to 58 years. The other SR included studies of women attending screening programs or with clinical suspicion of breast cancer.[20] The mean age of women in the studies from this SR ranged from 51 to 60 years.

One RCT from Norway involved all women attending a national screening program age 50 to 69 years of age (mean/median age not provided).[10] The other RCT was conducted in asymptomatic women age 40 to 49 in the United Kingdom who had previously undergone mammography and were deemed to be at moderate or high risk of breast cancer.[23] The mean age in this study was 44 years.

The cost-effectiveness study[24] was based on a study by Friedewald et al.[25] This was a retrospective analysis conducted in 13 institutions in the United States. Women in this study were attending screening examinations. The mean age in women receiving DBT plus DM was 56 years and the mean age in women receiving DM alone was 57 years.

In the Canadian guideline, the target population is women age 40 to 74 years of age not at increased risk of breast cancer.[4] The Brazilian guideline is targeted towards asymptomatic women and women with suspicious signs or symptoms.[26] The recommendations are also not directed at the female population at high risk of breast cancer.

Further details regarding eligible studies are in Appendix 2.

Interventions and Comparators

Three of the SRs[16,17,22] that solely investigated screening populations evaluated DBT in combination with DM the other[19] examined DBT alone or in combination with DM (though all eligible studies investigated DBT in combination with DM). The comparator in all four SRs was DM alone. The SR focusing on diagnosis examined DBT alone or combination with DM compared to DM alone.[21] One SR investigating both screening and diagnosis evaluated DBT alone or in combination with DM compared to DM alone.[18] The other SR of screening and diagnosis evaluated DBT in combination with DM versus DM.[20] In all the SRs, eligible studies had different processes for DBT with respect to reading (single or double) and the number of views (one or two). See Appendix 2 for further details of individual SRs.

Both guidelines made recommendations on DBT.

Outcomes

Three of the SRs focusing on screening evaluated cancer detection rate (CDR),[16,17,19] and two SRs evaluated recall rate (the proportion of screens requiring additional follow-up testing).[16,19] The SRs reporting on recall rate did not clarify this outcome further, though suggested that reduced recall rate would be a positive outcome as it reduces the burden of unnecessary testing.[16] One SR also evaluated sensitivity, specificity, and false positive rate.[19] One of the screening SRs measured interpretive efficiency, which was defined as the ratio of false positives to true positives.[22] Two of the SRs[17,19] reported outcomes according to cancer type and stage. The SR focusing on diagnosis measured sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, and diagnostic odds ratio. One SR of both screening and diagnosis measured CDR, recall rate, sensitivity, and specificity.[18] The other SR provided a narrative summary of sensitivity, false negative rate, validity, and precision.[20]

In one RCT, the primary outcome was the proportion of screen-detected breast cancer (CDR), and secondary outcomes were recall rate (proportion of women requiring follow-up assessment), and the positive predictive value (PPV) of recall and biopsy.[10] The primary outcome in the second RCT was recall rate, and the secondary outcome was CDR.[23]

The outcome in the economic evaluation[24] was cost per quality adjusted life year (QALY).

Please see Appendix 2 for additional details.

Summary of Critical Appraisal

Systematic reviews

All seven SRs had research questions incorporating PICO components. Two of the SRs[18,19] explicitly stated that a protocol was used, while the remaining SRs did not. The comprehensiveness of the search strategies varied across SRs. Three SRs had comprehensive search strategies,[16,19,20] three had partially comprehensive strategies (did not conduct grey literature search),[17,18,21] and one SR involved a Medline search and consultation with experts.[22] Study selection was conducted in duplicate in five SRs and was not clearly described in two.[16,22] Data extraction was conducted in duplicate in four SRs, by a single reviewer in one SR,[20] and it was unclear in two SRs.[19,22] Two SRs provided a list of excluded studies and reasons for exclusion,[19,21] while the others did not. Included studies were described in five of the studies, and poorly described in two.[16,21] Six of the SRs evaluated risk of bias (RoB) in included studies and one did not.[22] The SRs evaluating RoB in diagnostic studies all used QUADAS-2, which is an appropriate tool as it is specifically designed for critical appraisal in this context. None of the SRs reported the funding of the eligible studies. Meta-analysis was performed in six of the SRs and the methods used were appropriate in all six. However, three of the SRs did not describe the comparability of patient population and intervention (heterogeneity) with respect to pooling results,[16,19,21] and three did not examine the impact of RoB on results.[18,21,22] Heterogeneity of studies was narratively discussed in six of the SRs, and was not discussed in one.[22] One SR assessed for publication bias,[17] three stated that it was not possible or appropriate to assess for publication bias due to the number or design of studies,[16,19,21] while three SRs did not evaluate or comment on publication bias.[18,20,22] All SRs reported on conflict of interest of authors, though three[19,21,22] did not describe the funding source for their SR.

The SRs on screening contained little description of the populations in eligible studies, making it challenging to assess generalizability to a Canadian context. The screening SRs

generally investigated women attending population screening programs, with the mean/median age in most SRs around 55 to 60 years. This is broadly similar to the age distribution attending screening programs in Canada,[2] though it is unclear the extent to which the populations are reflective of the Canadian population in terms of other characteristics. One SR[18] included only women with dense breasts – the results of this SR may therefore be particularly applicable to this population, though not to the general population of women undergoing screening/diagnosis.

Further details of the critical appraisal are in Appendix 3.

Randomized controlled trials

Both RCTs were judged to be at low risk of bias with respect to patient selection, applicability, index test bias, and index test reliability. In both RCTs, it was unclear whether the person conducting the reference standard test was aware of the index test results. There was low risk of bias in both RCTs with respect to the applicability of the reference standard. However, both RCTs were at high risk of bias with respect to study flow, since not all trial participants received a reference test, which has the potential to bias accuracy results.[13] The cross-over RCT pooled data collected at different time points (one year apart) and it was unclear whether there was any change in condition during that period or how that might affect the findings. Further, in the crossover RCT conducted in the United Kingdom, it was unclear whether the person reading the second test was aware of the results of the first test.

One eligible RCT evaluated DBT for screening at a population level in Norway. These authors only described the age distributions of the participants, which was similar to the age distribution attending screening in Canada.[2] However, no other characteristics were described making it challenging to compare the study population to a Canadian context. The other RCT included a population of moderate to high risk women age 40 to 49, attending two clinics in the United Kingdom. As such, the results from this study may not be particularly generalizable to a screening population and to other populations for diagnosis.

Further details of the critical appraisal are in Appendix 3.

Economic evaluation

The economic evaluation[24] stated the research question and economic importance and relevance of the study. It also clearly outlined the form of economic evaluation, primary outcome measure, and model details. The authors described the effectiveness study on which the analysis was based though did not provide details of the study. Currency and price data were provided, while the discount rate and time horizon were also described; however, the discount rate was not justified. Further, the details of subjects from whom valuation was obtained were not given. The results were presented in disaggregated and aggregated form, the authors described sensitivity analyses, and conclusions were justified based on results. The authors noted they were unable to capture the cost of downstream workup for false positives, thus the benefit of a test having fewer false positives would be underestimated. The authors also assumed that all recalls involved radiologic biopsy rather than surgical biopsy (surgical biopsy is more expensive than radiologic biopsy according to the authors). They note that this may underestimate the benefit of a test having a reduced recall rate. The authors do not comment specifically on the effect of these possible underestimations on their results. However, they note that recall rates are lower for DBT plus DM, which suggests that cost-effectiveness may be underestimated by making these assumptions. The economic evaluation extrapolated downstream effects based on

detection differences and applied population cancer statistics (e.g. incidence, mortality) from the American Cancer Society. However, given the lack of direct clinical evidence on the effect of different testing strategies of health outcomes, it remains unclear what the true clinical effectiveness – and therefore cost-effectiveness – would be. Another limitation is that the authors did not include indirect costs in their model, which means their results do not account for factors such us lost work, transportation, child care, and other indirect costs. Finally, the authors assumed disutility values for testing and false-positives in breast cancer, and it is unclear the extent to which their assumptions are valid. This raises questions surrounding the validity of the model used.

The economic evaluation was based on American costs (Medicare data) and the performance of screening in an American retrospective study. As such, the generalizability to a Canadian context is not clear.

Further details of the critical appraisal are in Appendix 3.

Guidelines

Both guidelines clearly described their scope and purpose. The Canadian guideline[4] outlined stakeholder involvement, though no details about specific stakeholders were provided. The Brazilian guideline[26] did not describe the roles, professions, or contributions of guideline authors. The Canadian guideline was based on systematic searches for evidence and the methodology in developing recommendations was clearly outlined. It was not possible to assess the methodology of the Brazilian guideline as the methodology document was published in Portuguese. Both guidelines clearly described recommendations. The Canadian guideline contained details of resource implications and implementation, and clearly described competing interests of authors. The Brazilian guideline did not contain guidance for implementation, resource implications or auditing criteria, and conflict of interests of authors were not provided.

Further details of the critical appraisal are in Appendix 3.

Summary of Findings

Appendix 4 presents a table of the main study findings and authors' conclusions. There was overlap of primary studies included in the eligible SRs, particularly in the SRs from Marinovich et al.,[16] Yun et al.,[17] Phi et al.,[18] and Hodgson et al.[19] Appendix 5 presents a table outlining overlap of primary studies in the eligible SRs.

Clinical Effectiveness of DBT

No relevant evidence regarding the effectiveness (i.e. clinical benefits or harms) of DBT was identified; therefore, no summary can be provided.

Diagnostic accuracy of DBT

Cancer detection rate

Three SRs focusing on screening all reported that DBT plus DM improved CDR over DM alone.[16,17,19] Yun et al. reported on CDR for different types and stages of cancer.[17] These authors noted that DBT plus DM showed improved CDR over DM particularly in early stage, invasive cancer, while the benefit on detecting carcinoma in situ was uncertain. Hodgson et al. found that DBT plus DM improved detection rates for invasive cancer over DM alone, but that CDR for non-invasive cancer was no different between DBT plus DM and DM alone.[19]

One SR reported a subgroup of studies in the diagnostic setting in women with dense breasts.[18] These authors reported that DBT with or without DM improved CDR compared to DM alone (RR 1.12, 95% CI 1.01 to 1.24). This SR also reported a subgroup of studies in the screening setting in women with dense breasts. The authors found that DBT with or without DM improved CDR compared to DM alone in studies with two independent groups (one group of women received DBT+DM and the other group received DM) (RR 1.33, 95% CI 1.20 to 1.47) and where one group of women received both DBT plus DM and DM alone, and the results of each test were compared (RR 1.52, 95% CI 1.08 to 2.12).

The RCT from Hofvind et al. found no difference in the rate of screen-detected breast cancer for DBT plus DM compared to DM alone (RR 1.09, 95% CI 0.82 to 1.46).[10] The other RCT[23] found a detection rate of 0.51% (6/1175) in the DBT plus DM group and 0.43% (5/1170) in the DM alone group (no statistical testing was conducted for this outcome).

Recall rate

One SR of screening found the recall rate was lower for DBT plus DM compared to DM alone (difference = -2.2 per 1000 screens, 95% CI -3.0 to -1.4), though there was no difference in recall rate for studies involving paired samples (i.e. same individual receiving both tests)(difference = 0.5 per 1000 screens, 95% CI -0.1 to 1.2).[16] Another SR providing a narrative summary of screening found that DBT plus DM resulted in lower recall compared to DM alone, particularly for studies involving a single reading.[19]

One SR reported a subgroup of studies in the screening setting in women with dense breasts.[18] The authors found that DBT with or without DM lowered recall rate compared to DM alone in studies with two independent groups being screened (RR 0.72, 95% CI 0.64 to 0.80) but there was no difference where one group received both tests (RR 1.12, 95% CI 0.76 to 1.63).

One RCT[10] found a lower recall rate for DBT plus DM compared to DM alone (RR 0.78, 95% CI 0.69 to 0.88) while the other RCT[23] found no difference between techniques (2.7% for DBT plus DM versus 2.8% for DM alone, no p value provided).

Interpretive efficiency

One screening SR found that DBT plus DM improved interpretive efficiency compared to DM alone.[22] That is, the ratio of false positives to true positives was consistently lower for DBT plus DM.

False positive rate

One SR of screening found that the false positive rate was lower for DBT plus DM compared to DM alone, particularly for studies involving a single reading (whereas false positive rates were higher with DBT plus DM compared to DM alone with double reading).[19]

Sensitivity

One SR of screening studies identified a single study reporting sensitivity of DBT plus DM compared to DM alone with 12 month follow-up.[19] Sensitivity was better with DBT plus DM compared to DM alone; however, the authors stated they could not assess the potential benefit of DBT given limited data. One SR focusing on diagnosis compared sensitivity for DBT alone or in combination with DM to DM alone.[21] These authors pooled estimates for each technique separately and compared them narratively. The sensitivity for DBT alone

was 0.90 (95% CI 0.87 to 0.92) while the sensitivity for DM alone was 0.89 (95% CI 0.86 to 0.91).

Another SR reported a subgroup of studies conducted in a diagnostic setting in women with dense breasts.[18] These authors provided a narrative summary of evidence, suggesting the sensitivity of DBT with or without DM was higher than for DM alone in this setting.

Specificity

One SR of screening studies identified a single study reporting specificity of DBT plus DM compared to DM alone with 12 month follow-up. Specificity was better with DBT plus DM compared to DM alone; however, the authors stated they could not assess the potential benefit of DBT given limited data.[19] One SR focusing on diagnosis compared specificity for DBT alone or in combination with DM to DM alone.[21] These authors pooled estimates for each technique separately and compared them narratively. The specificity for DBT alone was 0.79 (95% CI 0.77 to 0.81) compared to 0.72 (95% CI 0.70 to 0.74) for DM alone.

Another SR reported a subgroup of studies conducted in a diagnostic setting in women with dense breasts.[18] These authors provided a narrative summary of evidence, suggesting the specificity of DBT with or without DM was no different than for DM alone in this setting.

Positive Likelihood Ratio

One SR focusing on diagnosis compared the positive likelihood ratio (PLR) for DBT alone or in combination with DM to DM alone.[21] These authors pooled estimates for each technique separately and compared them narratively. The PLR for DBT was 3.50 (95% CI 2.31 to 5.30) and for DM was 2.83 (95% CI 1.77 to 4.52).

Negative Likelihood Ratio

One SR focusing on diagnosis compared the negative likelihood ratio (NLR) was for DBT alone or in combination with DM to DM alone.[21] These authors pooled estimates for each technique separately and compared them narratively. The NLR for DBT was 0.15 (95% CI 0.06 to 0.36) and for DM 0.18 (95% CI 0.09 to 0.38).

Diagnostic odds ratio

One SR focusing on diagnosis compared the diagnostic odds ratio (DOR; odds of the test being positive if person has breast cancer relative to odds of test being positive if person does not have breast cancer) for DBT alone or in combination with DM to DM alone.[21] These authors pooled estimates for each technique separately and compared them narratively. The DOR for DBT was 26.0 (95% CI 8.70 to 78.0) and for DM 16.2 (95% CI 5.61 to 47.0).

Positive predictive value of recall

One RCT found the PPV of recall was higher for those receiving DBT plus DM compared to DM alone (PPV = 21.4% for DBT versus 15.2% for DM, *P* for difference = 0.011).[10]

Positive predictive value of biopsy

One RCT found the PPV of biopsy was no different for those receiving DBT plus DM compared to DM alone (PPV = 37.7% for DBT versus 32.1% for DM, p for difference = 0.18).[10]

Subgroup of women with dense breasts

One SR[18] reported on diagnosis and screening of women with dense breasts only. These authors provided a narrative summary of sensitivity and specificity, suggesting the sensitivity of DBT with or without DM was higher than for DM alone in the diagnostic setting, but there was no difference in specificity. The authors conducted meta-analysis for CDR and recall rate. They found that DBT with or without DM lowered recall rate compared to DM alone for screening in studies with two independent groups being screened (RR 0.72, 95% CI 0.64 to 0.80) but there was no difference where one group received both tests (RR 1.12, 95% CI 0.76 to 1.63). For diagnosis, the authors reported that DBT with or without DM improved CDR compared to DM alone (RR 1.12, 95% CI 1.01 to 1.24). For screening, DBT with or without DM improved CDR compared to DM alone in studies with two independent groups being screened (RR 1.33, 95% CI 1.20 to 1.47) and where one group received both tests (RR 1.52, 95% CI 1.08 to 2.12).

Cost-Effectiveness

The cost-effectiveness analysis[24] reported that DBT plus DM was cost-effective relative to DM alone. The authors reported the cost per QALY for DBT plus DM was $20 300 compared to DM alone. The willingness to pay threshold was $100 000/QALY. The cost per QALY was provided in 10 year age groups. In women age 40 to 49 the cost per QALY was $20 976; for women age 50 to 59 it was $49 725; for women age 60 to 69 it was $44 461; and for women age 70+ it was $82 500.

Guidelines

Both guidelines make strong recommendations against using DBT for screening. The Canadian guideline support this recommendation by stating there is no evidence surrounding the effect of DBT on patient-important outcomes, particularly noting “the recommendation is strong because these modalities would require the use of substantial and scarce health care resources when used for screening without evidence of benefit from their use” (p. E1444).[4] The Brazilian guidelines suggest that evidence is insufficient to determine the benefits and harms of DBT for screening.[26] These authors specifically note that it is unclear whether DBT provides benefit over DM alone, and whether benefits outweigh harms.

Limitations

The central limitation of the current body of evidence surrounding DBT is the absence of clinical evidence. The search for this report did not identify any studies examining clinical benefit/harm of DBT, and evidence informing the economic evaluation was extrapolated based on assumptions and population statistics. As such, it is unclear from this report whether the addition of DBT to DM (or DBT alone) would improve breast cancer mortality rates or reduce overdiagnosis compared to using DM alone for both screening and diagnosis. This has also been acknowledged in a recent review[8] on breast cancer screening and is reflected in guidelines identified for this report, which both recommend against using DBT due to lack of evidence surrounding its clinical benefit.[4,26]

While DBT appeared to generally improve detection and recall rates for screening, the eligible SRs did note sources of heterogeneity surrounding performance of DBT for these outcomes. This makes it challenging to draw conclusions surrounding the possible benefit of DBT with respect to CDR and recall rate for screening. For example, both the Yun[17] and Hodgson[19] SRs suggest that the benefit of DBT may depend on cancer type, noting benefit

for detection of invasive cancer but uncertainty for non-invasive cancer. Hodgson et al.[19] also suggested that recall rate for DBT may depend on the number of readers (these authors found discordant results with respect to recall rate when one reader was used compared to two). The Marinovich SR[16] suggested that possible benefit of DBT may be larger for biennial screening programs compared to annual screening programs. This SR also found that while recall rate was better for DBT and DM compared to DM alone when pooling all studies, in "paired" studies (where both modalities were used on same woman), there was no difference in recall rate. Phi et al.[18] also found no difference in recall rate for paired studies. As such, the possible benefit of DBT with respect to detection and recall may depend on the population (e.g. cancer type) and setting/protocol (frequency, number of readers, type of study [paired versus unpaired]).

The majority of evidence for screening reported on DBT in combination with DM (one SR reported on DBT with or without DM together), meaning there was no evidence on using DBT alone for screening. Both SRs that focused on diagnosis reported on DBT alone or in combination with DM, meaning the effect of each individual technique was unclear.

The effect of DBT on sensitivity and specificity is difficult to establish from the studies in this report. One SR evaluated the sensitivity or specificity of DBT in a screening setting, and only identified one study reporting these measures that had limited follow-up.[19] The authors therefore concluded it was not possible to assess DBT specificity and sensitivity. One SR that investigated sensitivity and specificity in the diagnostic setting pooled estimates for DBT and DM separately, and the authors did not conduct meta-analysis for the comparison between DBT and DM.[21] This SR comments on the difference between the two techniques but did not quantitatively compare them, making it difficult to evaluate benefit. Another SR in the diagnostic setting similarly provided a narrative summary on sensitivity and specificity but did not quantitatively evaluate a difference between the two.[18] Further, this SR included only women with dense breasts, making it challenging to compare the results of the two SRs that examined diagnosis.

Conclusions and Implications for Decision or Policy Making

This report identified seven SRs, two RCTs, one cost-effectiveness study, and two guidelines, on DBT alone or in combination with DM compared to DM alone of screening or diagnosis of breast cancer.

There was no evidence surrounding the effect of DBT on clinical outcomes of breast cancer, such as mortality or overdiagnosis. As such, the clinical benefit of using DBT alone or in combination with DM is unclear.

Evidence surrounding the effect of DBT in combination with DM compared to DM alone, on CDR and recall rate was conflicting. Eligible SRs generally concluded that DBT in combination with DM improved CDR and recall rate compared to DM alone for screening, but the potential benefit of DBT may vary according to cancer type, population, and screening setting/protocols. In contrast, both RCTs found no benefit of DBT in combination with DM on CDR, while one reported lower recall rate and the other found no benefit surrounding DBT on recall rate. One SR reported on sensitivity and specificity of DBT in a screening setting and concluded that due to limited follow-up data from only one study it was not possible to draw conclusions. Given conflicting evidence and uncertainties relating to heterogeneity, the available evidence makes it difficult to conclude whether DBT could be useful in a screening context.

Evidence on DBT in diagnosis was limited. For diagnosis, DBT alone or in combination with DM may improve sensitivity and specificity over DM alone; however, eligible studies did not quantitatively compare the two techniques and a narrative summary in one SR[18] suggested no difference between the two techniques with respect to specificity. In one SR of diagnosis for women with dense breasts, the CDR was improved with DBT alone or combination with DM compared to DM alone, but the evidence on CDR in a wider population is not clear. Limitations in evidence making it challenging to draw conclusions surrounding whether DBT would be useful for diagnosis of breast cancer.

One cost-effectiveness study conducted in the United States found that digital breast tomosynthesis in combination with digital mammography was cost-effective compared to digital mammography alone. However, the effectiveness estimates were based on assumptions regarding the downstream effects of different detection rates and American population statistics so the applicability to the Canadian setting is unclear.

The current state of evidence appears to be reflected in both Canadian and Brazilian guidelines, which both recommend against using DBT due to insufficient evidence surrounding its benefits and harms (notably for patient-important outcomes).[4,26] As such, long-term studies evaluating clinical outcomes, such as mortality, may be helpful in evaluating the potential benefit of DBT in both diagnosis and screening. It has also been suggested that future studies should include long-term follow up involving repeat screening, and assessment of interval cancer rates.[10,16,18]

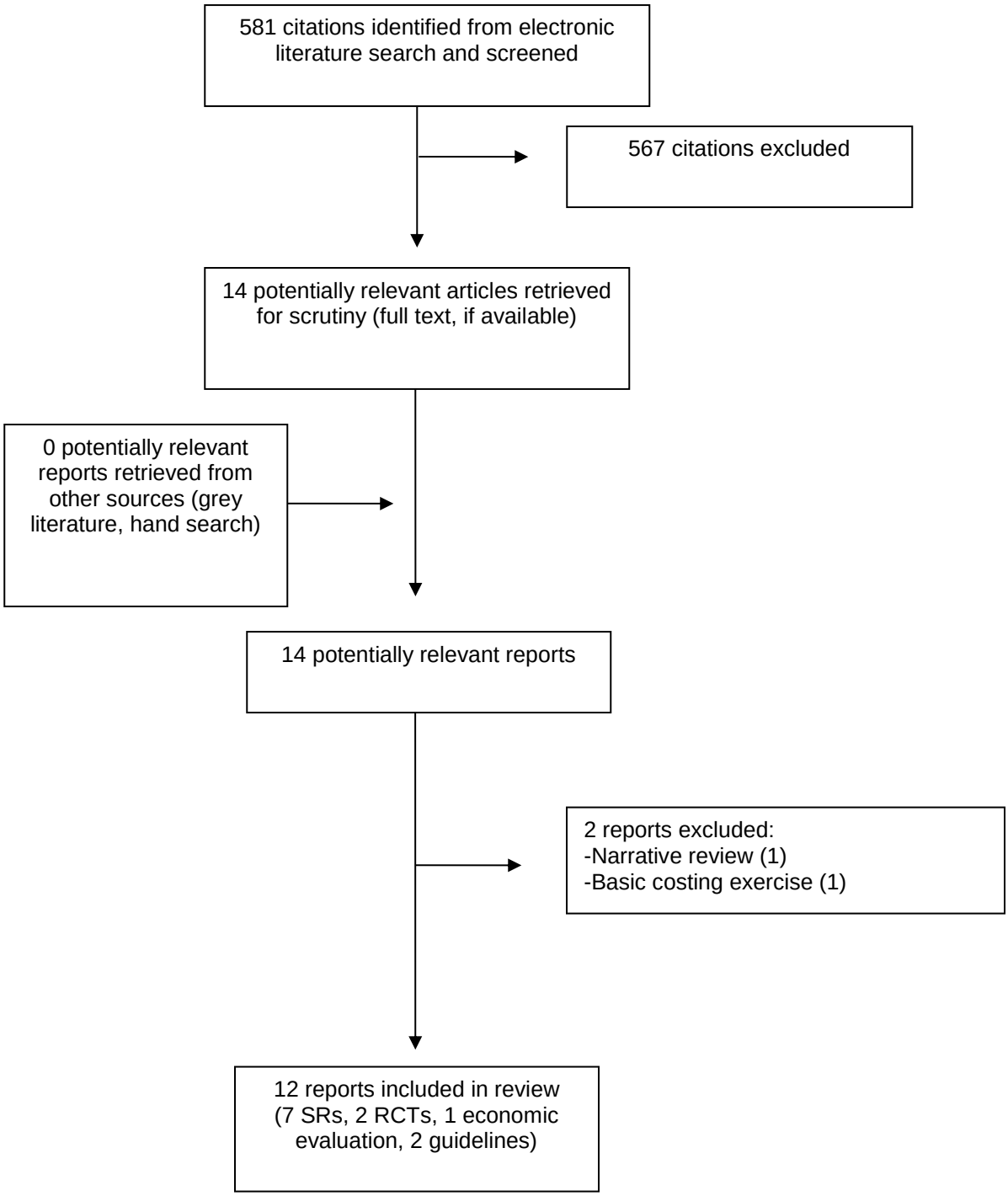
References

1. Canadian Cancer Society. Breast cancer statistics. 2019; https://www.cancer.ca:443/en/cancer-information/cancer-type/breast/statistics/?region=on. Accessed 2019 Oct 25.
2. Canadian Partnership Against Cancer. Breast cancer screening in Canada: monitoring and evaluation of quality indicators - Results report, January 2011 to December 2012. Toronto (ON): Canadian Partnership Against Cancer; 2017: https://s22457.pcdn.co/wp-content/uploads/2019/01/Breast-Cancer-Screen-Quality-Indicators-Report-2012-EN.pdf. Accessed 2019 Oct 25.
3. American Cancer Society. Breast cancer: facts & figures 2017-2018. Atlanta (GA): American Cancer Society, Inc; 2019: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf. Accessed 2019 Oct 25.
4. Klarenbach S, Sims-Jones N, Lewin G, et al. Recommendations on screening for breast cancer in women aged 40–74 years who are not at increased risk for breast cancer. *CMAJ.* Vol 1902018:E1441-1451.
5. Canadian Cancer Society. Get screened for breast cancer. 2019; http://www.cancer.ca/en/prevention-and-screening/reduce-cancer-risk/find-cancer-early/get-screened-for-breast-cancer/?region=on. Accessed 2019 Oct 25.
6. Canadian Cancer Society. Diagnosis of breast cancer. 2019; http://www.cancer.ca/en/cancer-information/cancer-type/breast/diagnosis/?region=on. Accessed 2019 Oct 25.
7. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review: A report jointly commissioned by Cancer Research UK and the Department of Health (England) October 2012. *Br J Cancer.* 2013;108(11):2205-2240.
8. Seely J, Alhassan T. Screening for breast cancer in 2018—what should we be doing today? *Curr Oncol.* 2018;25(Suppl 1):S115-124.
9. Elmore JG, Fletcher SW. Overdiagnosis in breast cancer screening: time to tackle an underappreciated harm. *Ann Intern Med.* 2012;156(7):536-537.
10. Hofvind S, Holen AS, Aase HS, et al. Two-view digital breast tomosynthesis versus digital mammography in a population-based breast cancer screening programme (To-Be): a randomised, controlled trial. *Lancet Oncol.* 2019;20(6):795-805.
11. Digital tomosynthesis for the screening and diagnosis of breast cancer: diagnostic accuracy, cost-effectiveness, and guidelines*. (CADTH Rapid response report: summary with critical appraisal).* Ottawa (ON): CADTH; 2019: https://www.cadth.ca/digital-tomosynthesis-screening-and-diagnosis-breast-cancer-diagnostic-accuracy-cost-effectiveness-0. Accessed 2019 Oct 25.
12. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ.* 2017;358:j4008. http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf. Accessed 2019 Oct 25.
13. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536.
14. Higgins JPT, Green S, editors. Figure 15.5.a: Drummond checklist (Drummond 1996). *Cochrane handbook for systematic reviews of interventions*. London (GB): The Cochrane Collaboration; 2011: http://handbook-5-1.cochrane.org/chapter_15/figure_15_5_a_drummond_checklist_drummond_1996.htm. Accessed 2019 Oct 25.
15. The AGREE Next Steps Consortium. The AGREE II Instrument. [Hamilton, ON]: AGREE Enterprise; 2017: https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf. Accessed 2019 Oct 25.
16. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast cancer screening using tomosynthesis or mammography: a meta-analysis of cancer detection and recall. *J Natl Cancer Inst.* 2018;110(9):942-949.
17. Yun SJ, Ryu CW, Rhee SJ, Ryu JK, Oh JY. Benefit of adding digital breast tomosynthesis to digital mammography for breast cancer screening focused on cancer characteristics: a meta-analysis. *Breast Cancer Res Treat.* 2017;164(3):557-569.
18. Phi XA, Tagliafico A, Houssami N, Greuter MJW, de Bock GH. Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts - a systematic review and meta-analysis. *BMC Cancer.* 2018;18(1):380.
19. Hodgson R, Heywang-Kobrunner SH, Harvey SC, et al. Systematic review of 3D mammography for breast cancer screening. *Breast.* 2016;27:52-61.
20. Garcia-Leon FJ, Llanos-Mendez A, Isabel-Gomez R. Digital tomosynthesis in breast cancer: a systematic review. *Radiologia.* 2015;57(4):333-343.
21. Lei J, Yang P, Zhang L, Wang Y, Yang K. Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis. *Eur Radiol.* 2014;24(3):595-602.
22. Svahn TM, Macaskill P, Houssami N. Radiologists' interpretive efficiency and variability in true- and false-positive detection when screen-reading with tomosynthesis (3D-mammography) relative to standard mammography in population screening. *Breast.* 2015;24(6):687-693.

23. Maxwell AJ, Michell M, Lim YY, et al. A randomised trial of screening with digital breast tomosynthesis plus conventional digital 2D mammography versus 2D mammography alone in younger higher risk women. *Eur J Radiol.* 2017;94:133-139.
24. Kalra VB, Wu X, Haas BM, Forman HP, Philpotts LE. Cost-effectiveness of tomosynthesis in annual screening mammography. *AJR Am J Roentgenol.* 2016;207(5):1152-1155.
25. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *Jama.* 2014;311(24):2499-2507.
26. Migowski A, Silva GAE, Dias MBK, Diz M, Sant'Ana DR, Nadanovsky P. Guidelines for early detection of breast cancer in Brazil. II - New national recommendations, main evidence, and controversies. *Cad Saude Publica.* 2018;34(6):e00074817.
27. Aujero MP, Gavenonis SC, Benjamin R, Zhang Z, Holt JS. Clinical performance of synthesized two-dimensional mammography combined with tomosynthesis in a large screening population. *Radiology.* 2017;283(1):70-76.
28. Bernardi D, Caumo F, Macaskill P, et al. Effect of integrating 3D-mammography (digital breast tomosynthesis) with 2D-mammography on radiologists' true-positive and false-positive detection in a population breast screening trial. *Eur J Cancer.* 2014;50(7):1232-1238.
29. Bernardi D, Macaskill P, Pellegrini M, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *Lancet Oncol.* 2016;17(8):1105-1113.
30. Brandt KR, Craig DA, Hoskins TL, et al. Can digital breast tomosynthesis replace conventional diagnostic mammography views for screening recalls without calcifications? A comparison study in a simulated clinical setting. *AJR Am J Roentgenol.* 2013;200(2):291-298.
31. Carbonaro LA, Di Leo G, Clauser P, et al. Impact on the recall rate of digital breast tomosynthesis as an adjunct to digital mammography in the screening setting. A double reading experience and review of the literature. *Eur J Radiol.* 2016;85(4):808-814.
32. Chae EY, Kim HH, Cha JH, Shin HJ, Choi WJ. Detection and characterization of breast lesions in a selective diagnostic population: diagnostic accuracy study for comparison between one-view digital breast tomosynthesis and two-view full-field digital mammography. *Br J Radiol.* 2016;89(1062):20150743.
33. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. *Lancet Oncol.* 2013;14(7):583-589.
34. Conant EF, Beaber EF, Sprague BL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography compared to digital mammography alone: a cohort study within the PROSPR consortium. *Breast Cancer Res Treat.* 2016;156(1):109-116.
35. Destounis S, Arieno A, Morgan R. Initial experience with combination digital breast tomosynthesis plus full field digital mammography or full field digital mammography alone in the screening environment. *J Clin Imaging Sci.* 2014;4:9.
36. Durand MA, Haas BM, Yao X, et al. Early clinical experience with digital breast tomosynthesis for screening mammography. *Radiology.* 2015;274(1):85-92.
37. Gennaro G, Toledano A, di Maggio C, et al. Digital breast tomosynthesis versus digital mammography: a clinical performance study. *Eur Radiol.* 2010;20(7):1545-1553.
38. Gilbert FJ, Tucker L, Gillan MG, et al. The TOMMY trial: a comparison of TOMosynthesis with digital MammographY in the UK NHS Breast Screening Programme--a multicentre retrospective reading study comparing the diagnostic performance of digital breast tomosynthesis and digital mammography with digital mammography alone. *Health Technol Assess.* 2015;19(4):i-xxv, 1-136.
39. Greenberg JS, Javitt MC, Katzen J, Michael S, Holland AE. Clinical performance metrics of 3D digital breast tomosynthesis compared with 2D digital mammography for breast cancer screening in community practice. *AJR Am J Roentgenol.* 2014;203(3):687-693.
40. Gur D, Abrams GS, Chough DM, et al. Digital breast tomosynthesis: observer performance study. *AJR Am J Roentgenol.* 2009;193(2):586-591.
41. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of tomosynthesis plus digital mammography and digital mammography alone for breast cancer screening. *Radiology.* 2013;269(3):694-700.
42. Houssami N, Macaskill P, Bernardi D, et al. Breast screening using 2D-mammography or integrating digital breast tomosynthesis (3D-mammography) for single-reading or double-reading--evidence to guide future screening strategies. *Eur J Cancer.* 2014;50(10):1799-1807.
43. Lang K, Andersson I, Rosso A, Tingberg A, Timberg P, Zackrisson S. 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. *Eur Radiol.* 2016;26(1):184-190.
44. Lourenco AP, Barry-Brooks M, Baird GL, Tuttle A, Mainiero MB. Changes in recall type and patient treatment following implementation of screening digital breast tomosynthesis. *Radiology.* 2015;274(2):337-342.
45. McCarthy AM, Kontos D, Synnestvedt M, et al. Screening outcomes following implementation of digital breast tomosynthesis in a general-population screening program. *J Natl Cancer Inst.* 2014;106(11).

46. McDonald ES, Oustimov A, Weinstein SP, Synnestvedt MB, Schnall M, Conant EF. Effectiveness of digital breast tomosynthesis compared with digital mammography: outcomes analysis from 3 years of breast cancer screening. *JAMA Oncol.* 2016;2(6):737-743.
47. Michell MJ, Iqbal A, Wasan RK, et al. A comparison of the accuracy of film-screen mammography, full-field digital mammography, and digital breast tomosynthesis. *Clin Radiol.* 2012;67(10):976-981.
48. Noroozian M, Hadjiiski L, Rahnama-Moghadam S, et al. Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology.* 2012;262(1):61-68.
49. Powell JL, Hawley JR, Lipari AM, Yildiz VO, Erdal BS, Carkaci S. Impact of the addition of digital breast tomosynthesis (DBT) to standard 2D digital screening mammography on the rates of patient recall, cancer detection, and recommendations for short-term follow-up. *Acad Radiol.* 2017;24(3):302-307.
50. Rafferty EA, Park JM, Philpotts LE, et al. Assessing radiologist performance using combined digital mammography and breast tomosynthesis compared with digital mammography alone: results of a multicenter, multireader trial. *Radiology.* 2013;266(1):104-113.
51. Rafferty EA, Durand MA, Conant EF, et al. Breast cancer screening using tomosynthesis and digital mammography in dense and nondense breasts. *Jama.* 2016;315(16):1784-1786.
52. Rose SL, Tidwell AL, Bujnoch LJ, Kushwaha AC, Nordmann AS, Sexton R Jr. Implementation of breast tomosynthesis in a routine screening practice: an observational study. *AJR Am J Roentgenol.* 2013;200(6):1401-1408.
53. Sharpe RE Jr, Venkataraman S, Phillips J, et al. Increased cancer detection rate and variations in the recall rate resulting from implementation of 3D digital breast tomosynthesis into a population-based screening program. *Radiology.* 2016;278(3):698-706.
54. Shin SU, Chang JM, Bae MS, et al. Comparative evaluation of average glandular dose and breast cancer detection between single-view digital breast tomosynthesis (DBT) plus single-view digital mammography (DM) and two-view DM: correlation with breast thickness and density. *Eur Radiol.* 2015;25(1):1-8.
55. Skaane P, Bandos AI, Gullien R, et al. Prospective trial comparing full-field digital mammography (FFDM) versus combined FFDM and tomosynthesis in a population-based screening programme using independent double reading with arbitration. *Eur Radiol.* 2013;23(8):2061-2071.
56. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. *Radiology.* 2013;267(1):47-56.
57. Skaane P, Bandos AI, Eben EB, et al. Two-view digital breast tomosynthesis screening with synthetically reconstructed projection images: comparison with digital breast tomosynthesis with full-field digital mammographic images. *Radiology.* 2014;271(3):655-663.
58. Starikov A, Drotman M, Hentel K, Katzen J, Min RJ, Arleo EK. 2D mammography, digital breast tomosynthesis, and ultrasound: which should be used for the different breast densities in breast cancer screening? *Clin Imaging.* 2016;40(1):68-71.
59. Svahn TM, Chakraborty DP, Ikeda D, et al. Breast tomosynthesis and digital mammography: a comparison of diagnostic accuracy. *Br J Radiol.* 2012;85(1019):e1074-1082.
60. Svane G, Azavedo E, Lindman K, et al. Clinical experience of photon counting breast tomosynthesis: comparison with traditional mammography. *Acta Radiol.* 2011;52(2):134-142.
61. Tagliafico A, Astengo D, Cavagnetto F, et al. One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *Eur Radiol.* 2012;22(3):539-544.
62. Teertstra HJ, Loo CE, van den Bosch MA, et al. Breast tomosynthesis in clinical practice: initial results. *Eur Radiol.* 2010;20(1):16-24.
63. Thibault F, Dromain C, Breucq C, et al. Digital breast tomosynthesis versus mammography and breast ultrasound: a multireader performance study. *Eur Radiol.* 2013;23(9):2441-2449.
64. Waldherr C, Cerny P, Altermatt HJ, et al. Value of one-view breast tomosynthesis versus two-view mammography in diagnostic workup of women with clinical signs and symptoms and in women recalled from screening. *AJR Am J Roentgenol.* 2013;200(1):226-231.
65. Wallis MG, Moa E, Zanca F, Leifland K, Danielsson M. Two-view and single-view tomosynthesis versus full-field digital mammography: high-resolution X-ray imaging observer study. *Radiology.* 2012;262(3):788-796.

Appendix 1: Selection of Included Studies



Appendix 2: Characteristics of Included Publications

Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses

First Author, Publication Year, Country	Objective, Study Designs, Numbers of Primary Studies Included, Databases, Search Date	Population Characteristics	Intervention and Comparator(s), Reference Standard	Outcome(s)
Screening				
Marinovich 2018, Australia[16]	**Objective** "summarize all the available evidence on cancer detection and recall for DBT vs DM screening and to assess heterogeneity in the evidence" **Studies** 17 studies total 13 retrospective nonrandomized studies (unpaired design; separate groups) 4 prospective nonrandomized studies (paired design; all participants had both techniques) **Databases** EMBASE, PREMEDLINE, DARE, HTA database, NHSEED, ACP Journal Club, Cochrane database, all via Ovid **Search date** July 2017	Asymptomatic women attending population breast cancer screening (n=1 009 790) Mean/median age 56.2 years in DBT group and 57.5 years in DM group Median proportion of high density (BIRADS 3 or 4): 46.6% in DBT group and 42.0% in DM group	**Intervention** DBT in combination with DM (n=350 810) **Comparator** DM (n=658 980) **Reference Standard** not described 16 studies used 2 view DBT, 1 study used 1 view DBT 13 studies involved a single read; 4 studies involved a double read	Detection rate Recall rate Authors separated results by paired design (all participants underwent both screening techniques) and unpaired design (separate groups underwent each technique) Length of follow up not described
Yun 2017, South Korea[17]	**Objective** "evaluate the benefit of adding DBT to FFDM compared to FFDM alone for breast cancer detection, with a focus on cancer characteristics"	Routine breast cancer screening Median age 54 to 59 years	**Intervention** DBT plus DM (n=112 624) **Comparator** DM alone (n=212 917)	Detection rate Results separated based on cancer characteristics (overall cancer, invasive cancer, carcinoma in situ)

Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses

First Author, Publication Year, Country	Objective, Study Designs, Numbers of Primary Studies Included, Databases, Search Date	Population Characteristics	Intervention and Comparator(s), Reference Standard	Outcome(s)
	Studies 11 studies total 3 prospective comparative studies 1 prospective observational study 7 retrospective observational studies **Databases** Pubmed, EMBASE, Cochrane Central **Search date** December 31, 2016	Rate of dense breast (definition not provided): 17 to 65%	**Reference standard** pathologic confirmation 6 studies used 2 view DBT; 2 studies presumed to use 2 view DBT; 2 studies used 1 view DBT; 1 study used MLO view DBT (number of reads not described)	
Hodgson 2016, United Kingdom[19]	**Objective** "examine the performance of DBT for breast cancer-screening" **Studies** 5 studies total (16 reports) 2 prospective comparative studies (fully paired) 3 retrospective reviews **Databases** MEDLINE, Embase, Cochrane Library, DARE, Cochrane Central, HTA database, SCI-EXPANDED, LILACS, Inspec, clinicaltrials.gov, EU clinical trials register, International Clinical Trials Registry Platform, NICE	Women participating in breast cancer screening program or undergoing opportunistic mammography screening; no history of breast cancer; no symptoms Mean/median age: 53 to 59 years	**Intervention** DBT alone or in combination with DM **Comparator** DM alone **Reference standard** excision histology or biopsy 2 prospective studies used 2 view DBT with independent double reader process 2 retrospective studies used single reader 1 retrospective study used double reader	Detection rate (overall and invasive cancer) Sensitivity Specificity False positive rate Recall rate Results separated by region (Europe versus USA) Follow up length not described

Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses

First Author, Publication Year, Country	Objective, Study Designs, Numbers of Primary Studies Included, Databases, Search Date	Population Characteristics	Intervention and Comparator(s), Reference Standard	Outcome(s)
	Search date October 2014 (studies excluded if conducted before 2008)			
Svahn 2015, Australia[22]	**Objective** "[examine] the effect of DBT on radiologists' interpretive efficiency in terms of true and false detection when screen-reading with DBT relative to standard mammography" **Studies** 3 studies total (4 publications) 2 prospective comparative studies 1 retrospective observational study **Database** Medline **Search date** July 2015	Population breast screening program participants Age range in studies: 48 to 71, 50 to 69, ≥18 years Number of screens in analysis: 43 148	**Intervention** DBT plus DM **Comparator** DM **Reference standard** As per eligible studies: histology, investigative imaging +/- histology, clinical record review, ultrasonography, magnetic resonance imaging 2 prospective studies used independent double read	Radiologist interpretive efficiency measured by false positive:true positive ratio (calculated for each individual study identified by the systematic review) Follow up length not described
Diagnosis				
Lei 2014, China[21]	**Objective** "provide a medical evidence basis for DBT in diagnosing breast lesions" **Studies** 7 studies total 5 retrospective studies (no further description)	Diagnosing breast lesions Population (n=2014) characteristics not described	**Intervention** DBT alone or in combination with DM **Comparator** DM alone **Reference standard** histological results (biopsy or surgery resection on follow-up)	Sensitivity Specificity Positive likelihood ratio Negative likelihood ratio Diagnostic odds ratio Length of follow up not described

Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses

First Author, Publication Year, Country	Objective, Study Designs, Numbers of Primary Studies Included, Databases, Search Date	Population Characteristics	Intervention and Comparator(s), Reference Standard	Outcome(s)
	2 prospective nonrandomized studies (comparing different techniques in same group of women) **Databases** Pubmed, EMBASE, Cochrane Library, Web of Science, Chinese Biomedical Literature Database, China Academic Journal database, Wanfang database **Search date** June 2013		No information provided about DBT	
Screening and diagnosis				
Phi 2018, The Netherlands[18]	**Objective** "systematically review the literature on the accuracy of DBT compared to DM in women with dense breasts" **Studies** 16 studies total 5 diagnostic studies (2 prospective cohort, 3 retrospective cohort) 8 screening studies using two independent groups (4 retrospective nonrandomized controlled studies, 4 retrospective studies with historical control group) 3 screening studies using one group (3 prospective studies)	Women older than 18 years who underwent breast imaging in a screening or diagnostic setting and were classified as having dense breasts on mammography (BIRADS 3 and 4/c and d); asymptomatic Diagnostic studies Mean/median age 49 to 58 Screening studies with two groups Mean/median age 56 to 58 Screening study with one group Mean/median age 56 to 58	**Intervention** DBT alone or in combination with DM (diagnostic n=2737, screening two groups n=115 838, screening one group n=6957) **Comparator** DM alone (diagnostic n=2737, screening two groups n=188 419, screening one group n=6957) Reference Standard: pathology Diagnostic studies: 2 studies used 2 view DBT and 3 studies used 1 view DBT; 3 studies used single read and 2 studies used double read	Detection rate Recall rate Sensitivity Specificity Separated results by diagnostic study, screening study using two separate groups, and screening study using one group Mean/median follow up ranged from 1 to 2 years where described; one study had a 6 month follow up

Table 2: Characteristics of Included Systematic Reviews and Meta-Analyses

First Author, Publication Year, Country	Objective, Study Designs, Numbers of Primary Studies Included, Databases, Search Date	Population Characteristics	Intervention and Comparator(s), Reference Standard	Outcome(s)
	Databases PubMed, Web of Science **Search date** May 2017		Screening studies with two groups: 6 used 2 view DBT and 2 did not described the number of views; 6 studies used single read and 2 studies did not describe Screening studies with one group: 2 studies used 2 view DBT and 1 study used 1 view DBT; 3 studies used double read	
Garcia-Leon 2015, Spain[20]	**Objective** "update the evidence available to establish effectiveness [of tomosynthesis], in terms of diagnostic validity and accuracy, in screening and breast cancer diagnosis" **Studies** 11 studies total 9 studies compared DBT to DM 8 prospective studies 1 case-control **Databases** Medline, EMBASE, Web of Science, Pubmed (annex 1) **Search date** February 2013	Women (n=2475) with clinical suspicion of breast cancer or who were included in screening programs Mean age: 51 to 60 years	**Intervention** DBT in combination with DM **Comparator** DM Authors did not explicitly describe intervention and comparator but narrative summary of results describes comparison between DBT+DM and DM alone Reference standard: biopsy or follow-up 4 studies used 2 view DBT; 2 studies used 1 or 2 view DBT; 3 studies used 1 view DBT; 2 studies did not describe	Narrative comparison of sensitivity, false negative rate, validity and precision Follow up ranged from 6 to 39 months; 5/11 studies had follow up of 12 months

Abbreviations: BIRADS = Breast Imaging Reporting and Data System; DBT = digital breast tomosynthesis; DM = digital mammography (2D mammography, full field digital mammography); MLO = mediolateral oblique

Table 3: Characteristics of Included Randomized Controlled Trials

First Author, Publication Year, Country	Study Design	Population Characteristics	Intervention and Comparator(s)	Clinical Outcomes, Length of Follow-Up
Hofvind 2019, Norway[10]	Parallel group randomized controlled trial	n=29 453 women age 50 to 69 years attending national screening program between January 14, 2016 and December 31, 2017	**Intervention** DBT including synthetic 2D mammograms (n=14 734) **Comparator** DM (n=14 719) **Reference standard** histologically verified ductal carcinoma in situ or invasive breast cancer, or both Both DBT and DM were 2 view; independent double read was used	**Primary outcome** proportion of screen-detected breast cancer **Secondary outcomes** proportion of recalls, positive predictive value of recall and biopsy **Follow-up** 12 months after recruitment period ended
Maxwell 2017, United Kingdom[23]	Randomized controlled crossover trial (authors describe primary study as parallel group randomized controlled trial)	Asymptomatic women (n=1227) age 40 to 49 who had previously undergone mammography and were at moderate or high risk of breast cancer as defined by NICE Mean age at recruitment = 44 years	**Intervention** DBT plus DM **Comparison** DM alone Crossover one year later Both DBT and DM were 2 view; independent double read was used	**Primary outcome** recall rate **Secondary outcome**s detection rate, individual reader recall rate

Abbreviations: DBT = digital breast tomosynthesis; DM = digital mammography (2D mammography, full field digital mammography)

Table 4: Characteristics of Included Economic Evaluation

First Author, Publication Year, Country	Type of Analysis, Time Horizon, Perspective	Decision Problem	Population Characteristics	Intervention and Comparator(s)	Approach	Clinical and Cost Data Used in Analysis	Main Assumptions
Kalra 2016, USA[24]	Cost-effectiveness analysis	Evaluate cost-effectiveness of annual screening with DBT	Overall population of women 40 to 79 years old with all breast densities;	**Intervention** DBT plus DM **Comparator** DM alone	Decision tree model based on data from Friedewald et al.[25]	**Diagnostic accuracy data** recall rates, invasive cancer detection rate,	No false negatives for tomosynthesis group (since used as

First Author, Publication Year, Country	Type of Analysis, Time Horizon, Perspective	Decision Problem	Population Characteristics	Intervention and Comparator(s)	Approach	Clinical and Cost Data Used in Analysis	Main Assumptions
	Time horizon lifetime **Perspective** federal payer	plus DM compared to DM alone in women of all breast densities age 40 to 79	reported overall and by each decade subgroup Population from Friedewald et al.[25] n= 454 850 examinations; mean patient age 57 years old for DM alone and 56.2 years for DBT plus DM			recall and biopsy probabilities **Costs** Medicare reimbursement values, cost of invasive and noninvasive breast cancer; did not include indirect costs such as lost work time or transportation	reference standard) Base case was woman 56 years or older presenting for annual screening No patients lost to follow-up Assumed disutility of testing and false positive Recall imaging based on cost of radiologic biopsies alone not on surgical biopsy; authors note that surgical biopsy is twice as expensive as radiologic

Abbreviations: DBT = digital breast tomosynthesis; DM = digital mammography (2D mammography, full field digital mammography)

Table 5: Characteristics of Included Guidelines

Intended Users, Target Population	Intervention and Practice Considered	Major Outcomes Considered	Evidence Collection, Selection, and Synthesis	Evidence Quality Assessment	Recommendations Development and Evaluation	Guideline Validation
Klarenbach 2018[4]						
Intended users Primary care providers **Target population** Women aged 40 to 74 years who are not at increased risk of breast cancer	Breast cancer screening with mammography	Breast cancer mortality All cause mortality Overdiagnosis False-positive results with ensuing biopsies	Review of reviews on outcomes of breast cancer screening for women aged 40 years and older not at increased risk of breast cancer; additional search in MEDLINE and Cochrane Library from October 2014 to January 2017 for primary studies; 3 systematic reviews were included Systematic review on women's values and preferences about screening from January 2000 to November 2016 (MEDLINE, Cochrane Library, CINAHL, PsycINFO, grey literature)	GRADE approach	GRADE approach Strong recommendation: "the task force is confident that the desirable effects of an intervention outweigh its undesirable effects (strong recommendation for an intervention) or that the undesirable effects of an intervention outweigh its desirable effects (strong recommendation against an intervention). A strong recommendation implies that most people will be best served by the recommended course of action." Conditional recommendation: "are those for which the desirable effects probably outweigh the undesirable effects (conditional recommendation in favour of an intervention) or undesirable effects probably outweigh the desirable effects (conditional recommendation against an intervention) but appreciable uncertainty exists."	Reviewed by content experts and stakeholders (details not provided)

Table 5: Characteristics of Included Guidelines

Intended Users, Target Population	Intervention and Practice Considered	Major Outcomes Considered	Evidence Collection, Selection, and Synthesis	Evidence Quality Assessment	Recommendations Development and Evaluation	Guideline Validation
Migowski 2018[26]						
Intended users Health professionals, health managers, general population **Target population** Asymptomatic women and women with suspicious signs or symptoms (not at high risk of breast cancer)	Breast cancer screening: mammography alone or in combination with self-examination, clinical breast examination, ultrasonography, MRI, breast tomosynthesis, thermography	Overall mortality Breast cancer mortality False-positives Overdiagnosis and overtreatment	Detailed methods provided in Portuguese Six systematic reviews plus additional three systematic reviews when updating evidence prior to publication (based on MEDLINE search to March 31, 2017)	Detailed methods provided in Portuguese	Detailed methods provided in Portuguese Strong recommendation against: “Most patients should NOT receive the intervention” and for patients “Most people, when well informed, would NOT want the intervention, only a minority would”	Not reported

Abbreviations: GRADE = Grading of Recommendations, Assessment, Development, and Evaluations

Appendix 3: Critical Appraisal of Included Publications

Table 6: Critical Appraisal of Systematic Reviews Using AMSTAR 2[12]

Item	Systematic Review						
	Marinovich 2018[16]	Phi 2018[18]	Yun 2017[17]	Hodgson 2016[19]	Garcia-Leon 2015[20]	Svahn 2015[22]	Lei 2014[21]
1. Did the research questions and inclusion criteria for the review include the components of PICO?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	No	Yes	No	Yes	No	No	No
3. Did the review authors explain their selection of the study designs for inclusion in the review?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
4. Did the review authors use a comprehensive literature search strategy?	Yes	Partial	Partial	Yes	Yes	No	Partial
5. Did the review authors perform study selection in duplicate?	No	Yes	Yes	Yes	Yes	Unclear	Yes
6. Did the review authors perform data extraction in duplicate?	Yes	Yes	Yes	No	No	Unclear	Yes
7. Did the review authors provide a list of excluded studies and justify the exclusions?	No	No	No	Yes	No	No	Yes
8. Did the review authors describe the included studies in adequate detail?	No	Yes	Yes	Yes	Yes	Yes	No
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	Yes	Yes	Yes	Yes	Yes	No	Yes
10. Did the review authors report on the sources of funding for the studies included in the review?	No	No	No	No	No	No	No
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	Yes	Yes	Yes	Yes	N/A	Yes	Yes
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on	No	No	Yes	No	N/A	No	No

Table 6: Critical Appraisal of Systematic Reviews Using AMSTAR 2[12]

Item	Systematic Review						
	Marinovich 2018[16]	Phi 2018[18]	Yun 2017[17]	Hodgson 2016[19]	Garcia-Leon 2015[20]	Svahn 2015[22]	Lei 2014[21]
the results of the meta-analysis or other evidence synthesis?							
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	Yes	No	Yes	Yes	Yes	No	No
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes	Yes	Yes	Yes	Yes	No	Yes
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	N/A	No	Yes	N/A	No	No	N/A
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Yes	Yes	Yes	No	Yes	No	No

Table 7: Critical Appraisal of Randomized Controlled Trials Using QADAS-2[13]

Item	Hofvind 2019[10]	Maxwell 2017[23]
Patient selection: risk of bias	Low	Low
Patient selection: concerns regarding applicability	Low	Low
Index test: risk of bias	Low	Low
Index test: concern regarding applicability	Low	Low
Reference standard: risk of bias	Unclear	Unclear
Reference standard: concerns regarding applicability	Low	Low
Flow and timing: risk of bias	High	High

Table 8: Critical Appraisal of Guidelines Using AGREE II[15]

Item	Guideline	
	Klarenbach 2018[4]	Migowski 2018[26]
Domain 1: Scope and Purpose		
1. The overall objective(s) of the guideline is (are) specifically described.	Yes	Yes
2. The health question(s) covered by the guideline is (are) specifically described.	Yes	Yes
3. The population (patients, public, etc.) to whom the guideline is meant to apply is specifically described.	Yes	Yes
Domain 2: Stakeholder Involvement		
4. The guideline development group includes individuals from all relevant professional groups.	Yes	No
5. The views and preferences of the target population (patients, public, etc.) have been sought.	Yes	Unclear
6. The target users of the guideline are clearly defined.	Yes	Yes
Domain 3: Rigour of Development		
7. Systematic methods were used to search for evidence.	Yes	Unclear
8. The criteria for selecting the evidence are clearly described.	Yes	Unclear
9. The strengths and limitations of the body of evidence are clearly described.	Yes	Partially
10. The methods for formulating the recommendations are clearly described.	Yes	Unclear
11. The health benefits, side effects, and risks have been considered in formulating the recommendations.	Yes	Yes
12. There is an explicit link between the recommendations and the supporting evidence.	Yes	Unclear
13. The guideline has been externally reviewed by experts prior to its publication.	Yes	Unclear
14. A procedure for updating the guideline is provided.	No	No
Domain 4: Clarity of Presentation		
15. The recommendations are specific and unambiguous.	Yes	Yes
16. The different options for management of the condition or health issue are clearly presented.	Yes	Yes
17. Key recommendations are easily identifiable.	Yes	Yes
Domain 5: Applicability		
18. The guideline describes facilitators and barriers to its application.	Yes	Partially
19. The guideline provides advice and/or tools on how the recommendations can be put into practice.	Yes	No

Table 8: Critical Appraisal of Guidelines Using AGREE II[15]

Item	Guideline	
	Klarenbach 2018[4]	Migowski 2018[26]
20. The potential resource implications of applying the recommendations have been considered.	Yes	No
21. The guideline presents monitoring and/or auditing criteria.	Yes	No
Domain 6: Editorial Independence		
22. The views of the funding body have not influenced the content of the guideline.	Yes	Unclear
23. Competing interests of guideline development group members have been recorded and addressed.	Yes	No

Table 9: Critical Appraisal of Economic Evaluation Using Drummond Checklist[14]

Item	Kalra 2016[24]
The research question is stated	Yes
The economic importance of the research question is stated.	Yes
The viewpoint(s) of the analysis are clearly stated and justified.	Yes
The rationale for choosing alternative programs or interventions compared is stated.	Yes
The alternatives being compared are clearly described	Yes
The form of economic evaluation used is stated	Yes
The choice of form of economic evaluation is justified in relation to the questions addressed.	Yes
The source(s) of effectiveness estimates used are stated.	Yes
Details of the design and results of effectiveness study are given (if based on a single study).	No
The primary outcome measure(s) for the economic evaluation are clearly stated.	Yes
Details of the subjects from whom valuations were obtained were given.	No
Quantities of resource use are reported separately from their unit costs.	No
Methods for the estimation of quantities and unit costs are described.	Yes
Currency and price data are recorded.	Yes
Details of currency of price adjustments for inflation or currency conversion are given	Unclear
Details of any model used are given.	Yes
The choice of model used and the key parameters on which it is based are justified.	Yes
Time horizon of costs and benefits is stated.	Yes
The discount rate(s) is stated.	Yes
The choice of discount rate(s) is justified.	No
Details of statistical tests and confidence intervals are given for stochastic data.	No

Table 9: Critical Appraisal of Economic Evaluation Using Drummond Checklist[14]

Item	Kalra 2016[24]
The approach to sensitivity analysis is given.	Yes
The choice of variables for sensitivity analysis is justified	Yes
The ranges over which the variables are varied are justified.	Yes
Relevant alternatives are compared.	Yes
Incremental analysis is reported.	Yes
Major outcomes are presented in a disaggregated as well as aggregated form.	Yes
The answer to the study question is given.	Yes
Conclusions follow from the data reported.	Yes
Conclusions are accompanied by the appropriate caveats	Yes

Appendix 4: Main Study Findings and Authors' Conclusions

Table 10: Summary of Findings Included Systematic Reviews and Meta-Analyses

Study	Outcome	Findings	Authors' Conclusion
Screening			
Marinovich 2018[16] **(all results from meta-analysis)**	CDR per 1000 (paired)	DBT+DM: 8.8 (95% CI 7.4 to 10.5) DM: 6.4 (95% CI 5.2 to 7.9) Difference: 2.4 (95% CI 1.9 to 2.9)	DBT improves initial screen detection measures (detection rate, recall rate); however, heterogeneity suggests effect of DBT may be different depending on setting (larger incremental difference in CDR for DBT with biennial versus annual screening programs)
	CDR per 1000 (unpaired)	DBT+DM: 5.7 (95% CI 5.3 to 6.0) DM: 4.5 (95% CI 4.1 to 5.0) Difference: 1.1 (95% CI 0.8 to 1.5)	
	CDR per 1000 (overall)	Difference 1.6 (95% CI 1.1 to 2.0) in favour of DBT+DM	
	Recall rate (%) (paired)	DBT+DM: 4.1 (95% CI 3.3 to 5.0) DM: 3.3 (95% CI 2.2 to 5.6) Difference: 0.5 (95% CI -0.1 to 1.2)	
	Recall rate (%) (unpaired)	DBT+DM: 8.0 (95% CI 6.5 to 9.8) DM: 11.3 (95% CI 9.6 to 13.3) Difference: -2.9 (95% CI -3.5 to -2.4)	
	Recall rate (%) (overall)	Difference: -2.2 (95% CI -3.0 to -1.4) in favour of DBT+DM	
Yun 2017[17] **(all results from meta-analysis)**	CDR for cancer	Higher detection rate for DBT+DM compared to DM alone (RR 1.29; 95% CI 1.16 to 1.43)	Benefit of DBT for screening was associated with T-/N- stage. DBT plus DM showed benefit over DM alone for detecting invasive cancer, particularly early invasive cancer (stage T1 or N0) and showed benefit regardless of histologic cancer grade or type; the benefit for detecting carcinoma in situ was uncertain Difficult to conclude the net clinical benefit of adding DBT
	CDR for invasive cancer	Higher detection rate for DBT+DM compared to DM alone (RR 1.33; 95% CI 1.17 to 1.51)	
	CDR for carcinoma in situ	Benefit for DBT+DM compared to DM alone uncertain (RR 1.20; 95% CI 0.94 to 1.52)	
	CDR for T1 (invasive cancer)	Higher detection rate for DBT+DM compared to DM alone (RR 1.39; 95% CI 1.14 to 1.70)	
	CDR for T2 or larger (invasive cancer)	Benefit for DBT+DM compared to DM alone uncertain (RR 1.39; 95% CI 0.90 to 2.16)	
	CDR for nodal negative (invasive cancer)	Higher detection rate for DBT+DM compared to DM alone (RR 1.45; 95% CI 1.21 to 1.74)	
	CDR for nodal positive (invasive cancer)	Benefit for DBT+DM compared to DM alone uncertain (RR 1.34; 95% CI 0.92 to 1.94)	
	CDR for grade I (invasive cancer)	Higher detection rate for DBT+DM compared to DM alone (RR 1.81; 95% CI 1.37 to 2.39)	
	CDR for grade II/III (invasive cancer)	Higher detection rate for DBT+DM compared to DM alone (RR 1.40; 95% CI 1.17 to 1.68)	
	CDR for ductal carcinoma (invasive cancer)	Higher detection rate for DBT+DM compared to DM alone (RR 1.44; 95% CI 1.19 to 1.74)	

Table 10: Summary of Findings Included Systematic Reviews and Meta-Analyses

Study	Outcome	Findings	Authors' Conclusion
	CDR for lobular carcinoma (invasive cancer)	Higher detection rate for DBT+DM compared to DM alone (RR 1.90; 95% CI 1.21 to 2.98)	
Hodgson 2016[19]	Difference in CDR per 1000 screens (meta-analysis)	Higher detection rate in favour of DBT+DM compared to DM alone (difference = 2.43; 95% CI 1.76 to 3.10)	European studies reported higher CDR for DBT+DM compared to DM alone; however, results for recall and false positive rates vary CDR and invasive cancer CDR higher using DBT+DM than with DM, but no difference for non-invasive cancer detection rates Limited follow up on interval cancer and only one study reporting data on sensitivity and specificity, so it was not possible to assess sensitivity and specificity "Evidence suggests that recall and false positive rates may be lower using DBT+FFDM, especially for single reader paradigms such as those common in the US." (p.60)
	Difference in CDR for invasive cancers per 1000 screens for European studies (meta-analysis)	Higher detection rate in favour of DBT+DM compared to DM alone (difference = 2.33; 95% CI 1.67 to 3.00)	
	Difference in CDR per 1000 screens for non-invasive cancer for European studies (meta-analysis)	No statistically significant difference in CDR between DBT+DM and DM alone (estimate not provided)	
	False positives (European studies; narrative summary)	One study found a lower false positive rate for DBT+DM compared to DM alone (difference per 1000 screens = -9.3; 95% CI -11.8 to -7.2) while another study found a lower false positive rate for a single reader but higher false positive rate after arbitration (difference per 1000 screens = 5.4; 95% CI 4.2 to 6.8)	
	Sensitivity (European studies; summary of 1 study)	One study provided follow-up data at 12 months DBT+DM: 90.8% (95% CI 80.7 to 96.5) DM: 60.0% (95% CI 47.1 to 72.0)	
	Specificity (European studies; summary of 1 study)	One study provided follow-up data at 12 months DBT+DM: 96.5% (95% CI 96.0 to 96.9) DM: 95.6% (95% CI 95.0 to 96.0)	
	Recall rate (European studies; narrative summary)	One study found lower recall rate for DBT+ DM compared to DM alone (difference per 1000 screens = -6.6; 95% CI -8.7 to -4.9) while another study a lower recall rate for a single reader but a higher recall rate after arbitration (difference per 1000 screens = 6.2; 95% CI 4.9 to 7.7)	
	Difference in CDR (US studies; narrative summary)	One large study found higher CDR in favour of DBT+DM compared to DM alone (difference per 1000 = 1.21; 95% CI 0.82 to 1.63); a smaller study found a non-statistically significant higher CDR in favour of DBT+DM (difference per 1000 = 1.91; 95% CI -6.43 to 10.25); another small study found non-significantly lower CDR in DBT+DM group (difference per 1000 = -0.76; 95% CI -2.5 to 0.97)	
	Difference in invasive cancer CDR (US studies; narrative summary)	Higher invasive CDR for DBT+DM versus DM alone for one large study (difference per 1000 screens = 1.20; 95% CI 0.80 to 1.60)	

Table 10: Summary of Findings Included Systematic Reviews and Meta-Analyses

Study	Outcome	Findings	Authors' Conclusion
		Smaller studies found lower CDR for DBT+ DM (difference per 1000 screens = - 0.94; 95% CI: -2.2 to 0.35) and no difference (estimate not provided)	
	False positives (US studies; narrative summary)	Proportion of false positives lower in DBT+DM group compared to DM alone in three studies examined: differences per 1000 = -17.4 (95% CI -19.2 to -15.6), -28.7 (95% CI -35.1 to -22.2), and -74.4 (95%CI -105.6 to -43.1)	
	Recall rate (US studies; narrative summary)	Lower recall in DBT+DM group compared to DM alone in all three studies: difference per 1000 = -16.2 (95% CI -18.0 to -14.5), -29.4 (95% CI -36.0 to -22.8), and -72.5 (95% CI -104.7 to -40.2)	
Svahn 2015[22]	Number of false positives per screen detected cancer (FP:TP – a lower FP:TP indicates better efficiency because fewer FPs are caused for each detected cancer)	FP:TP for DBT+DM versus DM: Study 1: 3.17 (95% CI 2.25 to 4.47) versus 5.96 (95% CI 4.08 to 8.72) Study 2: 7.07 (95% CI 4.99 to 10.02) versus 10.25 (95% CI 6.42 to 16.35) Study 3: 8.37 (95% CI 5.87 to 11.93) versus 20.84 (95% CI 13.95 to 31.12)	"Study-level pooled estimates of FP:TP ratios were substantially improved (i.e. lower FP:TP ratio) for all studies for DBT+DM relative DM alone" (p.691) "The majority of radiologists were more efficient screen-readers using DBT+DM (they had less FPs for each detected breast cancer) than using DM" (p.691)
Diagnosis			
Lei 2014[21]	Sensitivity (pooled estimate)	DBT+/-DM: 0.90 (95% CI 0.87 to 0.92) DM: 0.89 (95% CI 0.86 o 0.91)	"DBT had higher sensitivity and specificity for the diagnosis of benign and malignant lesions in breasts. These results illustrated the superior diagnostic accuracy of DBT relative to DM." (p. 601)
	Specificity (pooled estimate)	DBT+/-DM: 0.79 (95% CI 0.77 to 0.81) DM: 0.72 (95% CI 0.70 to 0.74)	
	Positive likelihood ratio (pooled estimate)	DBT+/-DM: 3.50 (95% CI 2.31 to 5.30) DM: 2.83 (95% CI 1.77 to 4.52)	
	Negative likelihood ratio (pooled estimate)	DBT+/-DM: 0.15 (95% CI 0.06 to 0.36) DM: 0.18 (95% CI 0.09 to 0.38)	
	Diagnostic odds ratio (pooled estimate)	DBT+/-DM: 26.0 (95% CI 8.70 to 78.0) DM: 16.2 (95% CI 5.61 to 47.0)	
Screening and Diagnosis			
Phi 2018[18]	Sensitivity of DBT with or without DM versus DM alone in women with dense breasts in **diagnostic setting**	DBT: 5 studies, ranged from 84% (95% CI 71 to 93) to 89% (95% CI 81 to 95) DM: 5 studies; ranged from 69% (95% CI 58 to 79) to 86% (95% CI 81 to 89)	"In both the screening and diagnostic settings, DBT improved CDR (versus DM) in women with dense breasts. In

Table 10: Summary of Findings Included Systematic Reviews and Meta-Analyses

Study	Outcome	Findings	Authors' Conclusion
	(narrative summary)		the diagnostic setting, using DBT with or without DM increased sensitivity but did not change specificity. There was a significant reduction in recall rate when using DBT with DM (versus DM) in retrospective screening studies comparing between two study groups, although heterogeneity across studies was relatively high. A small number of prospective studies conducted in organized screening programs did not show reduced recall from using DBT. Improved CDR and reduced recall rate from DBT may imply a more effective screening test or diagnostic work-up for women with dense breasts. However, the critical issue is that more studies with longer follow-up and more screening rounds are necessary to draw definite conclusions on whether this improvement in cancer detection has an impact on interval cancer rates and potentially on mortality." (p.8)
	Specificity of DBT with or without DM versus DM alone in women with dense breasts in **diagnostic setting** (narrative summary)	DBT: 5 studies; ranged from 72% (95% CI 68 to 72) to 93% (95% CI 89 to 96) DM: 5 studies; ranged from 57% (95% CI 55 to 59) to 94% (95% CI 91 to 97)	
	CDR using DBT with or without DM versus DM alone in women with dense breasts in **diagnostic setting** (pooled estimate from 3 studies)	Improved CDR with DBT (RR 1.12, 95% CI 1.01 to 1.24)	
	CDR for DBT with or without DM versus DM alone in **screening setting** (two groups; pooled estimate from 6 studies)	Improved CDR with DBT (RR 1.33, 95% CI 1.20 to 1.47)	
	Recall rate for DBT with or without DM versus DM alone in **screening setting** (two groups; pooled estimate from 7 studies)	Lower recall rate with DBT (RR 0.72, 95% CI 0.64 to 0.80)	
	CDR for DBT versus DM alone in **screening setting** (paired data; pooled estimate from 3 studies)	Improved CDR with DBT (RR 1.52, 95% CI 1.08 to 2.12)	
	Recall rate for DBT versus DM alone in **screening setting** (paired data; pooled estimate from 2 studies)	No difference in recall rate (RR 1.12, 95% CI 0.76 to 1.63)	
Garcia-Leon 2015[20]	Validity or precision (narrative summary)	9 studies measured validity or precision and found that DBT showed better results when combined with DM compared to DM alone (estimates not provided)	"The results for the diagnostic validity of tomosynthesis in the diagnosis of breast cancer were inconclusive and there were no results for its use in screening" (p.333)
	Global performance (measured by AUC; narrative summary)	Significantly better with DBT plus DM compared to DM alone (estimates not provided)	
	Sensitivity (narrative summary)	One study found that sensitivity of DBT was greater compared to DM (estimates not provided)	

Abbreviations: AUC = area under curve; CDR = cancer detection rate; CI = confidence interval; DBT = digital breast tomosynthesis; DM = digital mammography; FP = false positive; RR = relative risk; TP = true positive

Table 11: Summary of Findings Included Randomized Controlled Trials

Study	Outcome	Findings	Authors' Conclusion
Hofvind 2019[10]	Screen-detected breast cancer (%)	DBT+DM: 0.66% (95% CI 0.53 to 0.79) DM: 0.61% (95% CI 0.48 to 0.73) Difference: RR 1.09 (95% CI 0.82 to 1.46)	"This study indicated that digital breast tomosynthesis including synthetic 2D mammograms was not significantly different from standard digital mammography as a screening tool for the detection of breast cancer in a population-based screening programme." (p. 795)
	Recall rate (%)	DBT+DM: 3.1% (95% CI 2.8 to 3.4) DM: 4.0% (95% CI 3.7 to 4.3) Difference: RR 0.78 (95% CI 0.69 to 0.88)	
	Positive predictive value of recall (%)	DBT+DM: 21.4% (95% CI 17.6 to 25.2) DM: 15.2% (95% CI 12.3 to 18.2) (difference statistically significant, p = 0.011)	
	Positive predictive value of biopsy (%)	DBT+DM: 37.7% (95% CI 31.7 to 43.7) DM: 32.1% (95% CI 26.5 to 37.7) (difference not statistically significant, p=0.18)	
Maxwell 2017[23]	Detection rate	DBT+DM: 6/1175 (0.51%) DM: 5/1170 (0.43%)	"The addition of DBT to DM has no significant effect on the false positive recall rate in women in their forties with an increased risk of breast cancer undergoing incident screening." (p.138)
	Recall rate	DBT+DM: 2.7% DM: 2.8% (no significant difference, no p value provided)	
	False positive recall rate	DBT+DM: 2.2% DM: 2.4% (no significant difference, p=0.89)	

Abbreviations: CI = confidence interval; DBT = digital breast tomosynthesis; DM = digital mammography

Table 12: Summary of Findings of Included Economic Evaluation

Main Study Findings	Authors' Conclusion
Kalra 2016[24]	
Total discounted cost of DBT plus DM = $15 312 and 15.50 QALYs **Total discounted cost of DM alone** = $14 500 and 15.46 QALYs **ICER in total population** = $20 300/QALY (authors state this was below the WTP threshold of $100 000/QALY) **ICER in 10-year sub-groups:** Age 40 to 49: $20 976/QALY Age 50 to 59: $49 725/QALY Age 60 to 69: $44 641/QALY Age 70+: $82 500/QALY	"Our analysis found that the addition of annual screening DBT to DM beginning at the age of 40 years is cost-effective compared to DM alone." (p.1152)

Abbreviations: DBT = digital breast tomosynthesis; DM = digital mammography; ICER = incremental cost-effectiveness ratio; QALY = quality adjusted life year; WTP = willingness to pay threshold

Table 13: Summary of Recommendations in Included Guidelines

Recommendations	Strength of Evidence and Recommendations
Klarenbach 2018[4]	
"We recommend not using tomosynthesis to screen for breast cancer in women who are not at increased risk" (p.E1444) "No evidence was identified on the effect of breast cancer screening on outcomes important to patients. The recommendation is strong because these modalities would require the use of substantial and scarce health care resources when used for screening without evidence of benefit from their use." **(p.E1444)**	Strong recommendation (no evidence)
Migowski 2018[26]	
"Recommends against breast cancer screening with tomosynthesis, either alone or with mammography." (p.3) "Concerning tomosynthesis, two systematic reviews of studies on diagnostic accuracy identified heterogeneous results for sensitivity and specificity between the studies, besides validity problems. Thus, there is still no sufficient evidence to evaluate whether breast cancer screening with [this method] can bring some benefit and whether the possible benefits outweigh the harms" (p.6)	Strong recommendation (possible harms probably outweigh possible benefits)

Appendix 5: Overlap between Included Systematic Reviews

Table 14: Primary Study Overlap between Included Systematic Reviews

Primary Study Citation	Systematic Review Citation						
	Marinovich 2018[16]	Phi 2018[18]	Yun 2017[17]	Hodgson 2016[19]	Garcia-Leon 2015[20]	Svahn 2015[22]	Lei 2014[21]
Aujero 2017[27]	X						
Bernardi 2014[28]						X	
Bernardi 2016[29]	X	X	X				
Brandt 2013[30]					X		
Carbonaro 2016[31]		X					
Chae 2016[32]		X					
Ciatto 2013[33]	X	X	X	X		X	
Conant 2016[34]	X	X					
Destounis 2014[35]	X			X			
Durand 2014[36]	X		X				
Friedewald 2014[25]	X			X			
Gennaro 2010[37]					X		X
Gilbert 2015[38]		X					
Greenberg 2014[39]	X		X				
Gur 2009[40]							X
Haas 2013[41]	X	X					
Houssami 2014[42]				X			
Lang 2016[43]	X	X	X				
Lourenco 2015[44]	X		X	X			
McCarthy 2014[45]	X	X	X				

Table 14: Primary Study Overlap between Included Systematic Reviews

Primary Study Citation	Systematic Review Citation						
	Marinovich 2018[16]	Phi 2018[18]	Yun 2017[17]	Hodgson 2016[19]	Garcia-Leon 2015[20]	Svahn 2015[22]	Lei 2014[21]
McDonald 2016[46]		X					
Michell 2012[47]					X		X
Noroozian 2012[48]					X		
Powell 2017[49]	X		X				
Rafferty 2013[50]					X		
Rafferty 2016[51]		X					
Rose 2013[52]	X	X	X			X	
Sharpe 2016[53]	X		X				
Shin 2015[54]		X					
Skaane 2013[55]	X			X		X	
Skaane 2013b[56]			X	X			
Skaane 2014[57]				X			
Starikov 2016[58]	X						
Svahn 2012[59]					X		X
Svane 2011[60]							X
Tagliafico 2012[61]					X		
Teertstra 2010[62]							X
Thibault 2013[63]							X
Waldherr 2013[64]		X			X		
Wallis 2012[65]					X		