



Digital breast tomosynthesis

A literature review to inform BreastScreen Australia's position statement on the use of tomosynthesis in the assessment and diagnosis of breast cancer

Final report: 23 November 2018



ALLEN+CLARKE

Allen + Clarke has been
independently certified as
compliant with ISO9001:2015
Quality Management Systems



Document status:	Final
Version and date:	23 November 2018
Author(s):	Anna Gribble, Luke Scullion
Filing Location:	W:\Department of Health and Ageing Australia\BreastScreen Australia 2017\Position Statement on Tomosynthesis\Tomo in assessment\Report
Peer / technical review:	Dr Robyn Haisman-Welsh
Verification that QA changes made:	Anna Gribble
Proof read:	Anna Gribble
Formatting:	Salman Abbasnejad
Final QA check and approved for release:	Anna Gribble

CONTENTS

KEY TERMS	5
GUIDANCE ON HOW TO READ THIS REPORT	6
KEY FINDINGS	7
ASSESSMENT OF EVIDENCE SUMMARY	19
1. INTRODUCTION	20
1.1. About digital breast tomosynthesis	20
1.2. BreastScreen Australia's position statement on tomosynthesis	20
1.3. Purpose and scope of this literature review	21
1.4. Ongoing research	21
1.5. Imaging systems used in studies reported in this literature review	22
2. METHODOLOGY	23
2.1. Objectives	23
2.2. Research questions	23
2.3. Literature search	26
2.4. Limitations and interpretation	28
3. CAN DBT REPLACE DIGITAL MAMMOGRAPHY VIEWS IN THE BREAST CANCER ASSESSMENT CENTRE?	31
3.1. Diagnostic accuracy of DBT imaging compared to repeat FFDM or other DM imaging (clinical pathway not specified)	32
3.2. Diagnostic performance of DBT compared to DSCV	54
3.3. Inter-observer agreement increases with the use of DBT	60
3.4. Avoidance of unnecessary work-up for benign final outcome	64
3.5. Radiation dose	78
4. MAMMOGRAPHIC FINDINGS: IMPROVING CONSPICUITY	84
4.1. Compared to FFDM, DBT improves lesion conspicuity, making it easier for readers to determine whether a lesion is suspicious for malignancy	84
4.2. Architectural distortion	86
4.3. Focal asymmetry	96
4.4. Microcalcifications	101
4.5. Masses	108
5. CANCER CHARACTERISTICS	114
5.1. Cancer type	116
5.2. Tumour stage and grade	124
5.3. Biomarkers (receptor status)	130
6. DBT'S ROLE IN BIOPSY	133
6.1. There is some evidence that DBT-guided VAB improves lesion localisation, which contributes to faster biopsy procedures	136
6.2. There is promising evidence that DBT-guided VAB has a superior technical success rate	137

6.3.	DBT-guided VAB is effective in imaging mammographically occult lesions during biopsy	140
6.4.	One study reported that women were as satisfied with DBT-guided VAB compared to DM-guided VAB	141
7.	IMPLEMENTATION INTO THE BREAST CANCER ASSESSMENT CENTRE	142
7.1.	No research relevant to the Australian context was identified on the incremental costs associated with implementation	143
7.2.	Workflow could be improved with the implementation of DBT, but this depends on several practices	143
7.3.	Reader performance	148
7.4.	Reader training in DBT interpretation is required for reading screening and assessment settings	152
7.5.	DBT images are larger and IT requirements for storage and transmission need to be considered	153
8.	DIAGNOSTIC PERFORMANCE OF OTHER ADJUNCTIVE IMAGING IN THE ASSESSMENT OF LESIONS SUSPICIOUS FOR BREAST CANCER	154
8.1.	Adjunctive ultrasound	154
8.2.	MRI	159
	REFERENCES	165
	APPENDX A: POPULATIONS INCLUDED IN PRIMARY STUDIES	175
	APPENDIX B: QUALITY ASSESSMENT FOR INCLUDED STUDIES	176

KEY TERMS

1v-DBT	One-view DBT (not specified)
95% CI	95% confidence interval
AD	Architectural distortion
AUC	Area under curve
BIRADS	American College of Radiologists Breast Imaging Reporting and Data System
BSA	BreastScreen Australia
CADe	Computer aided detection
CC	Craniocaudal (view)
CCMM	Coned compression magnification mammography
CDR	Cancer detection rate
DBT	Two-view digital breast tomosynthesis (unless otherwise noted)
DBT _{CC}	One-view digital breast tomosynthesis (craniocaudal view)
DBT _{MLO}	One-view digital breast tomosynthesis (medio-lateral oblique view)
DCIS	Ductal carcinoma in situ
DM	Digital mammography
DM _{CC}	One-view digital mammography (craniocaudal view)
DSCV	Digital spot compression view
ER	Oestrogen receptor
FFDM	Full-field digital mammography (also known as two-view digital mammography)
FN	False negative
FNA	Fine needle aspiration
FOM	Figure of merit
HER-2	Human epidermal growth factor receptor 2
IDC	Invasive ductal carcinoma
ILC	Invasive lobular carcinoma
JAFROC	Jackknife free-response receiver operating characteristic
LCIS	Lobular carcinoma in situ
MGD	Mean glandular dose
mGy	Milligray
mSv	Millisievert
MLO	Mediolateral oblique (view)
MRI	Magnetic resonance imaging

MRM	Magnetic resonance mammography
OR	Odds ratio
PPV	Positive predictive value
PPV ₃	Positive predictive value for biopsy performed
PS-VAB	Prone stereotactic vacuum-assisted biopsy
R+D	Research and development
ROC	Receiver operating characteristic
RR	Relative risk
SFM	Screen-film mammography
s2DM	Synthesised two-view digital mammography
TP	True positive
VAB	Vacuum-assisted core biopsy

GUIDANCE ON HOW TO READ THIS REPORT

This report is a narrative literature review. It contains two main parts:

1. The *Key Findings* section provides a summary of the findings of this literature review presented by each research question. A GRADE assessment is also provided.
2. The main report provides detailed findings to inform the research questions. Because many of the studies and articles included in this paper covered multiple outcomes, we have presented the information by cancer characteristic or diagnostic performance metric rather than by study and article.

Appendix A includes the quality assessment tables (based on AMSTAR2 and the Scottish Intercollegiate Guidelines Network tools) for the systematic reviews and RCTs.

Throughout this literature review, there are also summaries from *Allen + Clarke's* literature review on the role of digital breast tomosynthesis (DBT) in screening asymptomatic women for breast cancer, where information from that literature review adds further depth to our understanding of the role of DBT in the assessment of lesions suspicious for breast cancer.

KEY FINDINGS

Background

The BreastScreen Australia (BSA) program detects the earliest signs of breast cancer in asymptomatic Australian women aged 50 to 74 years. All breast screening is completed using bilateral full-field digital mammography (FFDM). Certain imaging features (such as architectural distortion (AD), asymmetry, some microcalcification patterns, or mass) are predictive of malignancy. If suspicious imaging features are identified on a screening mammogram, the woman is recalled for further assessment. This assessment work-up is usually the same as that provided for a woman who has breast cancer symptoms. Work-up currently includes repeat or additional mammographic views, digital breast tomosynthesis (DBT is only available in some BSA program assessment centres), ultrasound, and/or (if required) biopsy. MRI may also be used later in the work-up process.

Purpose of this literature review

The Department of Health (Australia) contracted *Allen + Clarke* to undertake a literature review (not systematic review) on the use of DBT in the work-up of screen-detected abnormalities (i.e., its potential role in the assessment and diagnosis of breast cancer within a breast screening program). This literature review will support the Breast Screening Technical Reference Group's consideration of what (if any) changes are needed to the BSA position statement on DBT (including providing advice on the use of DBT in all BSA assessment centres). Further updates to the BSA position statement may be required as recruiting or active studies report interim or final findings.

Methodology

Allen + Clarke completed a systematic search of the OVID Medline, Embase, Proquest and SCOPUS databases as well as searches of health technology assessment, Cochrane and clinical trials databases covering a date range of 1 January 2010 to 31 May 2018. We used combinations of subject/index terms as appropriate to the search functionality of each database. Articles were included if they met pre-determined PICO(T/S) criteria. Studies reporting only on screening populations (i.e., DBT's role as a primary screening test) were excluded as were articles covered in *Allen + Clarke's* previous literature review on the role of DBT in screening asymptomatic women for breast cancer.

The evidence discussed in this literature review includes some larger population-based trials set in screening program assessment centres, multicentre studies and single institution observer studies using cancer-enriched datasets with retrospective designs (some of which may not have adequate bias control). Information about powering was generally lacking even if number of breast cancers in the cohort was described. No RCTs and few large paired/matched trials assessing the diagnostic performance of DBT were identified in our search. Included studies often did not clearly describe the clinical pathway in which DBT was used or the suspected lesions that DBT was being used to assess and/or studies did not clearly articulate the work-up views that the DBT modality was being compared to (for example, a study might only articulate DBT compared to DM views - unspecified). We have also reported results from studies investigating women recalled to assessment but note that many of the studies included both women recalled to assessment and those attending for work-up of a symptom of breast cancer (but the studies did not distinguish between results for the two groups).

Several breast screening populations (including some BSA state programs) and single institutions have implemented DBT into a clinical pathway for the assessment of suspicious mammographic findings identified at screening and/or symptomatic women. Also, there are five large prospective clinical trials which will report on the assessment performance and diagnosis accuracy of DBT compared to DM or ultrasound. Findings from these studies may provide further robust evidence and consistent findings about performance.

We found 117 relevant articles including four systematic reviews (two with meta-analysis of studies published before 2014), 14 narrative literature reviews, 96 studies, and three other papers (such as position statements). Limited Australian evidence was identified in this review: only one oral abstract was located. Primary studies already incorporated into systematic or narrative literature reviews were not further assessed unless additional material not described in these reviews was included. An overall summary table provides an indication of the strength of findings presented in this literature review.

DIAGNOSTIC PERFORMANCE OF DBT COMPARED TO DM IN THE BREAST CANCER ASSESSMENT CENTRE

Both DBT and DM are excellent diagnostic tests with DBT appearing to be diagnostically superior (as measured by AUC) but this finding is not definitive

There is a large body of evidence demonstrating benefit when DBT is used to assess findings suspicious for breast cancer (regardless of imaging modality, study design, sample or DBT unit). Both DM and DBT had very good to excellent overall diagnostic accuracy (based on a measurement of area under the receiver operating characteristic curve). Using ROC or JAFROC analysis to assess how accurately DBT and DM assess a woman's probability of having cancer, evidence (including pooled analysis from three systematic reviews), demonstrated a consistent (if varying) increase in DBT's performance. Increase in AUC measurement from combinations of/comparisons between DBT and DM from pooled analysis were reported at an AUC measurement of 0.8668 for one-view DBT compared to 0.8561 for DM. AUC measurement observed for FFDM + DBT ranged from 0.788 to 0.914 compared to 0.681 to 0.881. with FFDM. Median AUC measurement fell between 0.8-0.9 for both modalities, including results presented from the TOMMY trial with overall AUC measurement for DBT imaging protocols more likely to exceed 0.85. Overall, FFDM + DBT appears to provide at least equivalent assessment performance and diagnostic accuracy compared to FFDM, with some studies indicating superior (but not perfect) performance. Diagnostic performance with comparing a reconstructed 2D image (s2DM) shows promising performance results compared to FFDM.

With the use of DBT, less DSCV imaging is required

Data from six studies comparing DBT to DSCV reported very high AUC measurement and sensitivity for both imaging protocols. While DBT's equivalence or superiority was not demonstrated through statistical testing, the differences in AUC and sensitivity were small and usually favoured DBT. That is, AUC measurements for DBT ranged from 0.87 to 1 compared to 0.857 to 0.963 for DSCV. Reported sensitivity was also high: DBT's sensitivity was higher than or equivalent to DSCV, with DBT values ranging between 96.9% and 100% compared to 85.6% to 100%. Mixed results and wider values ranges were reported for specificity: studies reported either higher specificity with DSCV or equivalence. Data from the Maroondah study indicated that improved accuracy with DBT (compared to DSCV) would result in fewer biopsies and other imaging work-ups such as ultrasound. Reported diagnostic equivalence or superiority for DBT

compared to DSCV has already resulted in practice changes including in screening programs like the BSA and the NHSBSP (in locations where DBT-capable units are available) as well as in individual clinical practices.

DBT better depicts architectural distortion, focal asymmetry and masses

- ***Detection of architectural distortion appears to increase with DBT***

AD presentation is a subtle mammographic presentation that readers may have trouble detecting on DM. It can easily be missed, misinterpreted or misclassified. AD is a common finding on review of false-negative mammograms. It may indicate invasive breast cancer (particularly ILC and IDC). It can also indicate a benign finding. Correct identification of a benign lesion or malignancy is important so that cancers are not missed and that women do not undergo unnecessary assessment or diagnostic work-up.

Compared to FFDM, DBT improves AD conspicuity and increases the overall detection of subtle AD (whether related to a malignant or a benign structure). A systematic review using data from large screening trials reported that DBT detected an increase in cancers presenting as AD: 11.46 invasive cancers per 10,000 FFDM + DBT screening examinations compared to 3.63 invasive cancers detected per 10,000 FFDM screening examinations. This is likely due to the way DBT reduces masking of surrounding tissue and increases visibility of subtle but potentially important breast changes like AD. Improved conspicuity with DBT may provide readers with better information to consider whether AD presentation is suspicious for breast cancer.

- ***There are significant increases in inter-observer agreement for subtle mammographic presentations like AD***

Inter-observer agreement improved for subtle mammographic presentations like AD and asymmetries as did subjectively assessed reader confidence in detection of breast cancers with a subtle mammographic presentation. This is an important finding given the high PPV for malignancy that DBT-detected AD has (over 40%) and that AD is a primary mammographic finding characteristic of missed cancers. Such reduction in inter-observer variability and increases in inter-observer agreement are likely to translate to higher quality clinical care.

- ***FFDM-occult AD should be treated as suspicious for breast cancer***

Evidence suggests that AD presentation visible on DBT but occult on FFDM or ultrasound should be treated as suspicious for cancer given the high PPV₃ and the number of missed cancers identified by AD presentation on DBT alone; however, cancer presenting as AD is a difficult breast presentation to detect accurately.

- ***DBT detects clinically relevant breast cancers presenting as AD (especially those that are mammographically occult) but it also represents benign final outcome findings and may contribute to the unnecessary work-up of benign outcomes***

Research from symptomatic and asymptomatic populations highlights an issue with the unnecessary work-up of lesions with a final benign outcome and which may not have been clearly visible on FFDM imaging. Further research is needed to assess the sensitivity of DBT in distinguishing between malignant and benign structures if DBT is used in the assessment centre. AD presentation is a subtle mammographic finding that readers may have trouble detecting on DM. It can easily be missed, misinterpreted or misclassified. AD is a common finding on review of false-negative mammograms. AD may indicate invasive breast cancer (particularly invasive lobular carcinoma and invasive ductal carcinoma). It can also indicate a

benign finding. Correct identification of a benign lesion or malignancy is important so that cancers are not missed and that women do not undergo unnecessary assessment or diagnostic work-up.

- ***Use of DBT results in reclassification of focal asymmetries to suspicious masses OR more benign/normal, which may contribute to the unnecessary work-up of benign outcomes***

Asymmetries are a subtle mammographic finding which can be focal (likely to represent a breast cancer) or global (more likely to represent a benign breast structure). Increased lesion conspicuity and reduced tissue overlap with DBT compared to FFDM means that readers may be more confident in dismissing as benign, asymmetries which may look suspicious on DM.

All studies (almost all of which were in mixed study populations) reported improved conspicuity of asymmetry with DBT: two studies reported superior visibility in 75% of cases but did not describe any further implications. Other studies demonstrated that DBT or s2DM had superior sensitivity for asymmetry compared to FFDM (with incremental increases ranging from 6% to 27.5%). Asymmetry as a primary mammographic finding decreased with the use of DBT, with one larger study based on lesions recalled to assessment, indicating a decrease in recall with DBT of approximately 50%. Similarly, reported downgrading in BIRADS 3 classifications (i.e., inconclusive findings were resolved to normal or benign findings) was approximately 45% and were a result of improved confidence to dismiss an area as benign/normal or reclassification of the finding as a mass due to better visibility of the lesion margin and shape.

- ***DBT improves conspicuity of masses and enables better assessment of these for malignancy***

The shape, texture and appearance of tissue in and around a mass is indicative of whether a mass is benign or suspicious for malignancy. Imaging with DBT often results in reclassification of primary mammographic finding from a focal asymmetry to a mass. Improved lesion conspicuity with DBT reduces tissue overlap making it easier to determine mass margin by removing 'noise' and making it easier to determine between masses suspicious for malignancy and those that are benign. There is now sufficient evidence that DBT provides superior performance in terms of improving readers' view of mass margins/soft tissue lesions.

Mixed results are presented for microcalcifications: early studies suggested that DBT was an inferior imaging technique for microcalcifications, but later studies report greater equivalence

Specific patterns of microcalcifications are a key mammographic presentation for breast cancer, particularly DCIS. Calcifications in general have high x-ray attenuation properties meaning that microcalcifications suspicious for breast cancer are usually easy to detect on DM. Microcalcifications with patterning suspicious for breast cancer may be more difficult to detect on DBT due to the way that images are viewed (1mm slices compared to a larger 'slabbing' technique), or due to engineering issues in different DBT-capable units such as resolution.

Bearing in mind that DCIS is easily seen on FFDM, overall, studies report mixed results on whether DBT (either alone or as FFDM + DBT) has equivalent or inferior performance in terms of detecting microcalcifications compared to FFDM. There is some consensus that DBT alone may not be sufficient for the detection of cancers presenting with microcalcification as the primary mammographic finding. There is consensus that image quality is now equivalent to or

in some cases better than FFDM for microcalcifications. s2DM + view DBT images as a slab shows some promise in improving DBT's performance but further evidence of success is required.

When DBT is used, inter-observer agreement about whether a lesion is benign or malignant increases

Inter-observer agreement is an important measure of the overall accuracy of data collected to detect and evaluate breast lesions. Based on smaller retrospective observer studies in a range of diagnostic groups, current evidence suggests an increase in inter-observer agreement for BIRADS classification with the use of DBT compared to DM. This increase in agreement was observed in the following DBT reading protocols: FFDM compared to FFDM + DBT, FFDM compared to DBT alone, or FFDM compared to s2DM. Inter-observer agreement (as measured by kappa statistic) with DBT imaging increased in all studies that reported it, with the use of DBT increasing overall agreement from moderate to good or excellent. Reported increases were usually about 25%. FFDM + DBT appears to be a more reliable test for BIRADS agreement with kappa statistics exceeding 0.8 with much lower rates reported for FFDM (ranging from 0.58 to 0.873). The use of s2DM compared to FFDM also appears to improve inter-observer agreement, suggesting that lesion conspicuity is greater with the reduction in 'noise' available with the 3D reconstruction algorithm.

Use of DBT results in a decrease in the number of inconclusive results compared to DM views (i.e., DBT contributes to the avoidance of additional follow-up or further work-up for lesions with a benign final outcome)

In the BSA program, mammogram results are reported using the NCBCC Synoptic Breast Imaging Report; however, no research identified in this report used the Australian reporting system. Almost all studies used BIRADS classifications for reporting mammogram results.

Consistent findings were reported in all studies regardless of imaging protocol, study design, participants/study sub-population, or DBT unit used. Using DBT resulted in consistent changes to BIRADS lesion reporting. While not all results achieved statistical significance, all studies reported that:

- BIRADS 3 lesion reporting (i.e., inconclusive results) reduced with the use of DBT, implying greater reader confidence to classify a lesion as benign/normal or malignant and improving program accuracy by having:
 - fewer additional work-up views to determine malignancy or a benign structure
 - fewer cases of short interval follow-up (in programs/settings where this is the clinical response to an inconclusive mammogram result)
 - reduced biopsy for women with a benign final outcome, and
 - greater peace of mind and experience for women.
- BIRADS 1 or 2 lesion reporting increased meaning that more women did not have to undergo further imaging or follow-up (i.e., reduced workflow implications for assessment clinics), and/or

- BIRADS 4 and 5 lesion classifications increased in line with pathology-proven final malignant results (where available), suggesting improved predictive power for malignancy with DBT.

Increased reader confidence to dismiss or refer for biopsy is seen particularly in the reduction in indeterminate images (eg a change in the BIRADS 3 lesion classification).

Data from one prospective study and one large retrospective reader analysis in populations recalled to assessment looked at data pre- and post-implementation of DBT. They reported a decrease in the use of BIRADS 3 lesion reports and an increase in BIRADS 1 and 2 lesion reports with no decrease in sensitivity. This provides some assurance that mammographic findings can be more accurately classified when using DBT compared to than FFDM or DM alone. In these studies, fewer women were recalled to short-interval follow-up and were provided with a benign/normal finding with DBT. This is most likely due to improved lesion conspicuity and the availability of more information with which to assess margins that is available with DBT compared to FFDM. Given that BIRADS 3 lesions have a very low PPV for malignancy (typically less than 2%), increasing the accuracy of BIRADS 3 lesion classifications (without a commensurate decline in sensitivity) can improve diagnostic and health system performance. There is a commensurate reduction in unnecessary biopsy and fewer false positive results.

DBT detects more invasive disease compared to DM

DBT, alone or in combination with FFDM, detects significantly more invasive cancers compared to FFDM. Pooled analysis drawn from studies of women participating in screening as well as women recalled to assessment reported this result consistently. Key results included:

- a relative risk (RR) of 1.327 for the increase of invasive cancer with FFDM + DBT compared to FFDM
- significant increases in detection of invasive ductal carcinoma (RR: 1.437) and special type carcinomas like tubular, papillary, medullary, and mucinous carcinoma (an increase of more than eight percentage points), with DBT (in a range of imaging protocols) compared to FFDM, and
- DBT-detected but DM-occult cancers are more likely to be invasive cancers.

Studies in other populations (including symptomatic women) reported the same result.

- ***Invasive cancers may be detected at an earlier stage and different grade with DBT but there are some mixed results***

DBT appears to result in more accurate determination of lesion margin and therefore results in more accurate sizing of tumours. Pooled analysis and prospective studies from screening populations indicate that DBT detects a higher proportion of early stage breast cancers compared to FFDM. For example, pooled analysis from studies set in screening populations indicated increased detection of T1 cancer (RR 1.388) or an increase in T1 and T1N0 cancers. T1N0 cancers are smaller and are likely to not have spread to auxiliary lymph nodes making them clinically important as these cancers detected at screening are more likely to have a better prognosis.

DBT's benefit in detecting cancers T2 or larger was less certain (RR 1.391, 95% CI: 0.895, 2.163). No consistent differences in detection were reported for Grade II or III cancers.

Mixed results were reported for DCIS staging, with data from the TOMMY trial and two other retrospective studies indicating that DBT detected larger DCIS with a higher grade, which

supports some of the earlier findings that DBT may not accurately diagnosis cancers presenting as microcalcifications (although changes to reconstruction algorithms may have addressed this issue).

- ***It is unclear whether DBT detects more node negative disease compared to FFDM***

There were mixed results for the detection of node negative disease between the prospective and retrospective studies set in screening settings. The meta-analysis from screening/women recalled to assessment indicated an increase in the proportion of node negative disease detected with DBT, but other studies reported no difference in results by nodal status (or no change in grade or stage). This may reflect the screening imaging used (eg, FFDM + DBT or DCT + s2DM).

- ***DBT and FFDM perform similarly in hormone receptor status***

Regarding hormone receptor and HER-2 status of breast cancers, both modalities of (DBT and FFDM) imaging were comparable in their ability to detect both positive and negative status cancers.

Mixed results are reported for the detection of invasive lobular carcinoma

ILC can be difficult to detect with mammography because a single cell or single files of cells may not result in clear changes at mammographic presentation. Pooled analysis from prospective studies set in screening populations reported increased detection of invasive lobular carcinoma (RR: 1.901). This finding is consistent with studies that reported increased conspicuity of subtle mammographic presentations like asymmetry and AD, which can be indicative of ILC. Two prospective studies reporting on ILC did not detect statistically significant increases in ILC detection compared to FFDM. One retrospective study that specifically focused on ILC detection reported an increase in detection with DBT.

For women recalled to assessment or symptomatic women, DBT does not appear to detect more in situ carcinoma (particularly DCIS) compared to DM

Pooled analysis results and results from prospective studies indicated that DBT did not appear to detect more in situ carcinoma including DCIS compared to DBT (RR 1.198, 95% CI: 0.942, 1.524). In the studies reported in this literature review, FFDM detected proportionally more DCIS cases across a range of studies in diagnostic populations (including populations of women recalled to assessment). DBT detected more cancers than FFDM overall but it did not preferentially detect more non-invasive cancers, which may never become clinically significant.

Some screening programs and clinics have already implemented DBT into assessment

While there are limitations in the evidence base (small sample sizes, retrospective designs, cancer-enriched samples, use of DBT-capable prototypes, etc.), findings presented in the primary studies have been sufficient to drive changes in clinical and screening program practices (both internationally and within the BSA program) in relation to the assessment of lesions suspicious for breast cancer. DBT's superiority in terms of AUC measurement and sensitivity is often cited as the reason for change.

DBT-guided VAB is an emerging technique that offering promising improvements in technical success and procedure performance speed

- ***DBT provides clear x, y, z coordinates from first imaging and improves localisation which contributes to faster biopsy time***

DBT is a helpful tool for triangulating lesions that are seen on only one view on either FFDM or DBT. This alleviates the need for additional mammographic projections for lesion localisation or confirmation. Precise targeting of a lesion's location can be calculated from one DBT projection, which is a substantial improvement over DM. Evidence from smaller retrospective studies in clinical environments suggests that DBT-guided VAB supports faster biopsy because coordinates data is available more quickly and, in one study, DBT-guided VAB was up to 43% faster than DM-guided VAB.

- ***Emerging evidence indicates that DBT-guided VAB has a superior technical success rate compared to DM-guided biopsy, resulting in reduction in avoidable harm to women from repeat biopsy***

Evidence from two small retrospective studies in women with lesions suspicious for breast cancer indicate that DBT-guided VAB has a superior technical success rate compared to DM-guided VAB. Study results show that DBT-guided VAB was successful in 100% of biopsies (including those presenting with subtle mammographic findings like AD), with stereotactic-guided VAB having a slightly lower success rate.

There is moderate evidence suggesting that lesions found using DBT imaging that are occult from DM or ultrasound images should be considered suspicious for malignancy and should undergo biopsy. These findings have positive implications for workflow in terms of fewer repeat biopsies and faster procedure times.

- ***DBT-guided VAB is emerging as a promising technique but more research is needed***

DBT-guided VAB is an emerging technique. The literature comparing stereotactic-guided VAB to DBT-guided VAB is limited to a small number of little studies; however, in current literature suggests that DBT has promising potential for use in guiding biopsy (especially for lesions that are FFDM or sonographically occult). These findings are based on less than 400 biopsies, all of which investigated the use of DBT-guided VAB. In addition, this literature review found no evidence discussing avoidance of benign biopsy where calcification was the main presentation.

- ***One study reported that women were satisfied with DBT-guided VAB***

Patient satisfaction with DBT-guided VAB reported similar overall patient satisfaction as that reported for stereotactic VAB.

No research relevant to the Australian context was identified on the incremental costs associated with implementation

No research papers identified in this literature review discussed the incremental costs associated with implementing DBT into an assessment centre.

Assessment centre workflow could be improved by the implementation of DBT, but this depends on several practices

Implementing DBT affects workflow but the overall impact is driven by when in the screening and assessment process DBT is implemented.

If used as a screening tool, DBT may result in reduced recall to assessment imaging (or reduced further diagnostic work-up) due to improved lesion conspicuity and more accurate initial reading of screening mammograms. This could result in fewer inconclusive screening results and decrease the necessity of short-term follow-up studies; however, this is likely to be of limited benefit in the BSA program as short-term follow-up is not a preferred management technique (i.e., the program aims to resolve all results to normal/benign or malignant). DBT when used in screening may result in fewer false positive recalls, resulting in reduced work-up of benign final outcome lesions, which could offset the lengthier image interpretation times associated with DBT; however, this literature review did not explore the full pathway of possibilities given its focus on the role of DBT in the breast cancer assessment centre.

If DBT can acquire sufficient diagnostic data in a single compression (if using FFDM + DBT), it potentially results in:

- faster diagnostic work-up because women do not have to have further mammographic work-up (i.e., reduced positioning time compared to additional DM views, fewer positioning errors and reduced need for repeat imaging, and reduced time needed to move between different types of imaging equipment)
- shorter overall additional time under compression for women, and
- fewer total diagnostic imaging studies or lesion localisation views.

DBT's impact on the need for other work-up views (including ultrasound and MRI) remains unclear, with some early studies indicating a small decrease in the use of ultrasound following implementation of DBT into the assessment centre, but others not reporting a significant decrease (i.e., ultrasound is still required, especially for the work-up of masses).

Finally, DBT has a higher PPV₃ indicating better targeted biopsy and, when DBT is used with VAB, faster biopsy times.

- ***Reading times for DBT are longer compared to DM images but this time could be offset by a reduction in the number of women being recalled to assessment***

DBT requires larger data storage compared to DM modalities. Images take much longer to read compared to DM simply because there are many more images to scroll through; however, overall, clinics might find that the extra time is not much of an issue if fewer women are recalled to assessment.

All readers improve their diagnostic accuracy when using DBT but there is some evidence to suggest that less experienced readers improve more when using DBT compared to more experienced readers

Consistent evidence (mostly based on small numbers of cases and readers but including a large sub-study from the TOMMY trial) indicates that all readers improve their performance with DBT, but less experienced readers improve more when using either FFDM + DBT or DBT_{MLO/CC} compared to FFDM or DSCV. Three retrospective studies reported that greater gains tend to

accrue to less experienced readers. While different definitions are used to describe “more” or “less” experience (including prior DBT experience and prior DM experience by volume or time), the trend is the same. Possible reasons for the increase in less experienced readers may be due to improved lesion conspicuity with DBT, making it easier to detect abnormalities presenting as subtle mammographic findings (especially AD and asymmetry) which may be missed on FFDM by less experienced readers. Only one retrospective study reported mixed results with some readers performing better with FFDM compared to FFDM + DBT; however, several important bias’ within the study were not accounted for.

- ***More research is needed to validate these findings***

More studies are needed to validate the influence of prior DBT/DM experience and confirm that it is a combination of prior experience with DM that drives diagnostic accuracy gains (i.e., readers transition to DBT smoothly with only a small amount of additional training) and the influence of improved lesion conspicuity for those with less experience who can then better visualise subtle findings, improving detection.

Research efforts include developing effective new imaging protocols to reduce a woman’s lifetime exposure to mammography-based assessment of suspicious lesions

- ***Per view, the MGD is higher for FFDM + DBT compared to FFDM alone or additional spot views***

Overall reported real-world MGD for DBT compared to FFDM views were similar to that reported in *Allen + Clarke’s* previous literature review: per view, DBT alone has a similar MGD compared to DM but the combined mode (FFDM + DBT) exposes women to almost double the radiation dose. Therefore, investigating ways to reduce the dual acquisition dose remains an important area of study.

- ***DBT’s MGD per view should be considered in the context of lifetime radiation dose from all mammography (screening + additional assessment imaging)***

Intuitively, there is likely to be considerable variation in the average difference between MGD for an imaging work-up involving FFDM + DBT compared to FFDM + DSCV because women are likely to require a different number of DSCV views to collect all the necessary information to inform diagnosis and treatment planning. Less variation in dose would be expected with DBT as the necessary information is likely to be available in one or two views. Further research is needed to confirm this. While the DBT imaging dose is higher than FFDM, women are less likely to be recalled for further assessment and therefore experience both fewer instances of supplementary DSCV and, if DBT provides sufficient diagnostic information, avoided overall DM work-up across her lifetime.

- ***Efforts to optimise diagnostic accuracy and reduce radiation dose continue***

Research to determine the lowest possible radiation dose needed to acquire satisfactory images continues. There is some emerging evidence that DM_{CC} + DBT_{MLO} results in minimal increase in MGD with a good improvement in lesion detection; however, further work to test the diagnostic accuracy of this imaging protocol is required, including considering how it might interface with s2DM, which also results in decreased radiation dose but performance that is equivalent to FFDM + DBT.

DIAGNOSTIC PERFORMANCE OF OTHER ADJUNCTIVE IMAGING IN THE ASSESSMENT OF LESIONS SUSPICIOUS FOR BREAST CANCER

Ultrasound

This literature review identified limited evidence discussing the role of adjunctive DBT (i.e., FFDM + DBT) compared to adjunctive ultrasound (FFDM + ultrasound) in the work-up of screen-detected breast cancers. Identified literature included populations undergoing a range of screening and/or diagnostic pathways and a range of different imaging view combinations.

- ***DM + DBT does not have inferior diagnostic accuracy compared to DM + ultrasound, but there is limited evidence about DBT's role in a work-up pathway compared to ultrasound and there is no suggestion that DBT should replace ultrasound views during work-up***

Two prospective studies reported on different imaging protocols and study populations (either comparing DM + DBT to DM + ultrasound or comparing a range of techniques) making it challenging to compare findings. Both prospective studies were powered to determine the non-inferiority of adjunctive DBT compared to adjunctive ultrasound and, despite some mixed findings on AUC measurement, non-inferiority was confirmed. Retrospective observational studies reported similar results to the prospective studies: adjunctive DBT was not inferior to adjunctive ultrasound but the superiority of adjunctive DBT has yet to be demonstrated. This evidence does not suggest that FFDM + DBT should replace FFDM + ultrasound in the work-up of suspicious breast cancers. It is not possible to assess whether there are some situations in which FFDM + DBT would be preferred over FFDM + ultrasound.

MRI

Six studies investigated adjunctive DBT compared to adjunctive MRI, with two prospective studies commenting on whether DBT could provide sufficient additional diagnostic information to result in a reduction in use of MRI in breast cancer assessment.

- ***In symptomatic women and mixed populations, adjunctive DBT is not inferior to adjunctive MRI for the assessment of lesions suspicious for breast cancer but there is no evidence to suggest that DBT should replace MRI in the work-up of suspicious lesions***

No statistically significant gain in sensitivity with adjunctive MRI was reported if a lesion had been imaged with FFDM + DBT + ultrasound: sensitivity with FFDM + ultrasound + MRI was 98.8% compared to 97.7% for FFDM + DBT + ultrasound. Overall, consistent findings were reported in the retrospective observational studies: adjunctive MRI has better diagnostic accuracy (as measured by AUC measurement) than adjunctive DBT but MRI had a lower specificity.

ASSESSMENT OF EVIDENCE SUMMARY

Table 1 (below) presents a short summary of the quality of evidence presented in the four systematic reviews informing this literature review (Phi et al., 2018; Yun et al., 2017; Garcia-Léon et al., 2015; Lei et al., 2014).

Table 1: Assessment of evidence for FFDM + DBT

Outcomes	Participants Studies	Quality of evidence	Overall results
Area under the receiver operating characteristic curve (AUC measurement)	Imaged with DM: 120,551 Imaged with DBT: 60,192 Three systematic reviews (34 studies)	⊕⊕ Low	DBT increases AUC measurement but there is significant heterogeneity in the primary studies and some controlled confounding
Sensitivity	Imaged with DM: 120,551 Imaged with DBT: 60,192 Three systematic reviews (34 studies)	⊕⊕ Low	Increased sensitivity with DBT, improved diagnostic OR with DBT and increased pooled positive likelihood ratio but there is significant heterogeneity in the primary studies and some controlled confounding
Specificity	Imaged with DM: 120,551 Imaged with DBT: 60,192 Three systematic reviews (34 studies)	⊕⊕ Low	Mixed results reported for specificity but there is significant heterogeneity in the primary studies and some controlled confounding
PPV/NPV	No SR reported on this	No data	No data in systematic reviews or RCTs
BIRADS agreement	No SR reported on this	No data	No data in systematic reviews or RCTs
Cancer type	Imaged with FFDM: 175,825 Imaged with FFDM + DBT: 75,532 Paired trials: 37,085	⊕⊕⊕ Moderate	Increased detection of invasive cancer with FFDM + DBT compared to FFDM No increase in detection of DCIS with FFDM + DBT compared to FFDM
Tumour staging and grading	Imaged with FFDM: 23,220 Imaged with FFDM + DBT: 18,090 Paired trials: 29,585 One SR (five studies)	⊕⊕⊕ Moderate	Increased detection of T1 cancer with FFDM + DBT No increase in detection of >T2 cancers with FFDM + DBT compared to FFDM Increased detection of invasive N0 cancer with FFDM + DBT compared to FFDM; no increased detection of ≥N1 with FFDM + DBT
Histological grade	Imaged with FFDM: 52,530 Imaged with FFDM + DBT: 30,442 Paired trials: 29,793 One SR (five studies)	⊕⊕⊕ Moderate	Increased detection of grade I, II and III cancers with FFDM + DBT compared to FFDM Increased detection of IDC and ILC with FFDM + DBT compared to DBT

1. INTRODUCTION

1.1. About digital breast tomosynthesis

Digital breast tomosynthesis (DBT) (also known as breast tomosynthesis, mammographic tomosynthesis or three dimensional/3D mammography/pseudo 3D mammography) is an imaging technology that can be used to detect and assess lesions suspicious for breast cancer and to diagnose breast cancer. DBT records between 11 and 25 low-dose images of a compressed breast depending on the imaging system used.¹ These images are reconstructed in 1mm (or more) parallel slices (or stacks) to form a three-dimensional image of the breast. Radiologists (or other readers) then analyse these images to determine the presence of suspected abnormalities or to further investigate an area identified as suspicious on 2D digital mammography (DM) or (in some cases) ultrasound. The thin cross-sectional images created by DBT reduce the masking effects of breast tissue overlap, which can improve margin visibility and increase lesion conspicuity. This increases sensitivity (especially for women with dense/non-fatty breasts) as abnormalities are easier to see. It may reduce specificity slightly if benign lesions (including those that are FFDM-occult) are then identified for the first time with DBT.

Breast imaging technologies like DBT already play a role in the assessment of lesions suspicious for breast cancer. When used in the assessment of a lesion suspicious for breast cancer, DBT provides additional imaging information to the breast care team (including information about exact location of the abnormality in the breast, lesion characteristics and cancer type). This information supports accurate diagnosis and prognosis assessment, informs the need for other diagnostic procedures (such as biopsy), and supports treatment planning including surgical management. DBT can reduce the need for further procedures (such as additional work-up views or biopsy) by faster dismissal of benign lesions.

1.2. BreastScreen Australia's position statement on tomosynthesis

In 2014, the Community Care and Population Health Principal Committee of the Australian Health Ministers' Advisory Council endorsed BSA's position statement on DBT. This position statement was based on a literature review completed in 2009 (Department of Health and Ageing, 2009) and other papers published between 2009-13. It includes a statement about the role of DBT in the assessment of breast cancer:

"There is evidence that tomosynthesis can be of benefit in an assessment setting".²

Further, the BSA position statement on DBT states that it:

"has the potential to decrease the number of women who are recalled for further tests (reduce recall rates) and possibly increase the detection of breast cancer (improve sensitivity)."; however, the balance between

"relative harms and benefits to well women of radiation dose, and the cost, efficiency and effectiveness of using this technology are as yet unclear".

¹ Hologic's Dimensions system takes 15 projections taken over approximately 4 seconds. Other CE mark or FDA-approved systems use 9 or 25 projections taken over 3 to 25 seconds (Sechopoulos, 2013).

² Cited papers included Public Health England's current position statement on use of tomosynthesis (DBT) in the NHS Breast Screening Programme (2013), Zuley et al. (2013) and Michell et al. (2012).

The Standing Committee on Screening concluded that full-field digital mammography (FFDM) remained the most effective population screening technology for breast cancer. It provided no further comment about the role of DBT in the assessment or diagnosis of breast cancer.

Since publication of BSA's position statement, the evidence base underpinning DBT's role in relation to other assessment of suspicious lesions or diagnostic work-up has developed further. DBT is now used in many breast cancer assessment clinics. In Australia, private radiology providers and some (but not all) BSA assessment centres use DBT to obtain additional information about suspicious areas on a screening mammogram or for women presenting with symptoms. DBT may also be used as part of a work-up to confirm breast cancer (for example, supporting biopsy). An interim Medicare rebate for DBT is now available and can be claimed for women with a past occurrence of breast cancer (or a family history) or who have symptoms or indications of cancer (including from a positive screening mammogram completed as FFDM).

1.3. Purpose and scope of this literature review

The Department of Health (Australia) engaged Allen and Clarke Policy and Regulatory Specialists Limited (*Allen + Clarke*) to support an update of the BSA's position statement on DBT (if the evidence suggests that this is required). *Allen + Clarke* was contracted to:

- complete a literature review on the role of DBT in the work-up of screen-detected abnormalities (i.e., its potential role in the assessment and diagnosis of breast cancer within a breast screening program) (this report), and
- prepare any updates to the BSA's position statement on DBT if considered necessary by the Breast Screening Technical Reference Group.

We wanted to know if DBT is a more sensitive and specific imaging technology that can contribute to reduced work-up for benign final outcome compared to DM views including repeat FFDM, digital spot-compression or magnification views (DSCV), or adjunctive ultrasound and adjunctive MRI. This includes considering evidence assessing whether DBT should be the preferred method of assessment for women with a suspected malignancy identified by screening, and any incremental costs or safety considerations associated with the implementation of DBT as an imaging tool in assessment or diagnosis. Initially, the answers to these questions will support the Breast Screening Technical Reference Group's consideration of what updates (if any) are needed to BSA's position statement on DBT.

The literature review is not a systematic review. No original meta-analysis or other pooled analysis was completed. We searched for the comparators most likely to be used instead of (or alongside) DBT in the assessment or diagnosis of breast cancer. This literature therefore looks at DBT compared to imaging techniques like repeat FFDM, other supplementary DM views including DSCV, and other adjunctive techniques like MRI or ultrasound.

The Department has commissioned a separate literature review on the role of DBT in screening. Short summaries of relevant sections of that report are included in this literature review for completeness and ease of access to information where it is relevant to the evidence presented in this report.

1.4. Ongoing research

Our review of www.clinicaltrials.gov (completed on 8 July 2018) identified six recruiting or active studies investigating the role of DBT in the assessment or diagnosis of breast cancer. Some of these studies are also investigating DBT's role as a screening tool but will report on

cancer characteristics (including the Tomosynthesis Trial in Bergen study³, the Tomosynthesis Mammographic Imaging Screening Trial⁴, and the PROSPECTS trial in the United Kingdom⁵). Other studies specifically related to the assessment or diagnostic performance of DBT compared to DM, ultrasound or MRI are:

- Tomosynthesis or Ultrasound in Mammographically-Negative Breasts (TOMUS)⁶
- Evaluation of Tomosynthesis for Characterization and the Management of Breast Lesions (ETOLE)⁷, and
- A small (100 participants) clinical trial in Canada to compare the performance of DBT to FFDM in detecting and characterizing suspicious findings in women having breast biopsies.

Further, research from the Maroonah trial is yet to be published (although an oral abstract is available, Lockie et al., 2014, refers). Additional updates to the BSA position statement may be required as these studies report interim or final findings.

1.5. Imaging systems used in studies reported in this literature review

DBT systems may differ by imaging geometry, angular range, number of projections, scan duration, radiation dose, acquisition method, detector technology, and reconstruction algorithms. We assessed the imaging systems used in all primary studies included in this literature review. Approximately two-thirds of studies were completed on Hologic's Selenia Dimensions unit. Other studies used Siemens Mammomat Inspirations, GE's SenoClaire or Senographe, or prototype units (either earlier beta units or new units under development by KERI or FUJI). This literature review also includes studies reporting on DBT + s2DM: all these studies used Hologic's Selenia Dimensions system and C-view 2D software for digital mammography image acquisition (which became available following FDA-approval in 2013).

³ This prospective cohort study (29,453) women compares DBT + s2DM to DM as a screening tool for women aged 50-69 years participating in a population-based screening program. Relevant study outcomes include prognostic and predictive tumour characteristics. This study began in January 2016 and primary study completion is set for January 2020.

⁴ The TMIST trial is a randomised trial of up to 164,946 Canadian women to compare diagnostic accuracy of screening for breast cancer with FFDM + DBT compared to FFDM alone. Key relevant outcomes include prevalence of breast cancer subtypes detected with DBT compared to FFDM alone, clinical characteristics of detected cancers, BIRADS assessment, biopsy rates and biomarker correlation. The primary lead-in study completion date is November 2018. The final study completion date is 2030.

⁵ The PROSPECTS RCT has a proposed sample of 100,000 women to investigate the cost-effectiveness of breast cancer screening using FFDM + DBT compared to DBT + s2DM. It aims to demonstrate that DBT is not inferior to FFDM + DBT. Relevant study outcomes include benign biopsy rates at diagnostic assessment and surgery. The RCT will happen over seven years from 2018 with initial results to be presented within 18-24 months.

⁶ The TOMUS trial is an Italian observational cohort study of 8000 women to investigate equivalence between DBT and ultrasound to determine if ultrasound could be substituted by DBT. The primary completion date is September 2018.

⁷ The ETOLE clinical trial will evaluate if the BIRADS classification obtained by DBT + s2DM is superior to that obtained by FFDM in specificity but not inferior in sensitivity. It involves 2000 participants and will be completed in May 2021.

2. METHODOLOGY

Summary

- This literature review provides an overview of research about the effectiveness and safety of DBT as an imaging tool in the work-up of suspicious lesions or diagnosis of breast cancer (detected through screening mammography or in symptomatic women). It is not a systematic review. We have provided statements about the quality of the evidence included in this review. No primary research or pooled analysis was undertaken.
- The following databases were searched in June 2018: EMBASE, Ovid Medline, CINAHL, ProQuest and Scopus. The following websites were reviewed: clinicaltrials.gov, the Cochrane database, NICE, INAHTA, and the UK NHSBPS.
- All returned citations and abstracts were assessed for relevance to the research questions and inclusion criteria. The same criteria were used to review the full-text and bibliographies of all articles proposed for inclusion. The methodologies of all included studies were critically appraised using the AMSTAR 2 tool or SIGN criteria.
- A total of 117 articles met the inclusion criteria and are discussed in this review.
- Limited Australian evidence was identified in this review: only one oral abstract was located with further publications from the Maroondah trial expected soon.

2.1. Objectives

This literature review explores if DBT is a more sensitive, specific and safer imaging technology for the work-up of screen-detected abnormalities (i.e., its potential role in the assessment and diagnosis of breast cancer within a breast screening program) compared to digital mammographic views, including repeat FFDM, DSCV, or adjunctive ultrasound or adjunctive MRI. It describes evidence on:

- overall diagnostic accuracy, accuracy of lesion classification and accuracy by mammographic finding, inter-reader agreement, and diagnostic performance using DBT by reader experience
- radiation dose compared to work-up with other DM views,
- characteristics of cancers detected with DBT compared to FFDM views
- DBT's role in biopsy, and
- implementation of DBT as an assessment tool.

A systematic review with pooled analysis was not performed.

2.2. Research questions

This literature review explored three questions. One question related to DBT's performance in the assessment and diagnosis of breast cancer; one question related to DBT's performance in relation to biopsy; and one question related to implementation considerations.

2.2.1. DBT's performance in the assessment and diagnosis of breast cancer

Research question 1 was:

For women aged over 40 years who undergo assessment for breast cancer (either after being recalled to assessment or symptomatic):

- is DBT (alone or with s2DM) a more sensitive, specific and safer technology for assessing suspected breast cancer compared to digital mammographic views, MRI or ultrasound imaging?
- is DBT (alone or with s2DM) a more sensitive, specific and safer technology for diagnosing breast cancer compared to digital mammographic views, MRI or ultrasound imaging?

The PICO(T/S) criteria underpinning research question 1 are described in *Table 2* (below).

Table 2: PICO(T/S) criteria for questions relating to DBT's performance in the assessment and diagnosis of breast cancer

Criterion	Description
Population	Women aged over 40 years who undergo assessment for breast cancer after being recalled to assessment because of a suspicious area in the breast detected through a screening mammogram or women with symptoms of breast cancer <i>NB women with more dense/non-fatty breasts are a key sub-population group</i>
Intervention	DBT (either alone or when combined with s2DM)
Comparators	FFDM Digital mammography alone (including spot-compression or magnification views) Hand-held ultrasound MRI
Outcomes	Lesion conspicuity Tumour margin assessment Type/sub-type of breast lesions and localization Microcalcifications, architectural distortions Morphology and dimension Radiation dose Reduction in ultrasound use Reduction in unnecessary biopsy Observer/radiologist diagnostic performance Women's anxiety
Study types	Systematic reviews, randomised controlled trials, observational studies

Research question 2 was:

For women aged over 40 years who undergo assessment for breast cancer after being recalled to assessment or who are symptomatic, does DBT-guided biopsy:

- improve lesion localisation compared to biopsy guided by digital mammography alone or ultrasound imaging?
- result in fewer procedural complications compared to biopsy guided by digital mammography or ultrasound imaging?

The PICO(T/S) criteria underpinning research question 2 are described in *Table 3* (below).

Table 3: PICO(T/S) criteria for questions relating to DBT's role in biopsy

Criterion	Description
Population	Women aged over 40 years who undergo assessment for breast cancer after being recalled to assessment because of a suspicious area in the breast detected through a screening mammogram or women with symptoms <i>NB women with more dense/non-fatty breasts are a key sub-population group</i>
Intervention	DBT (either alone or when combined with s2DM) as a perioperative guide
Comparators	Mammographic stereotactic guided biopsy Biopsy guided by hand-held ultrasound
Outcomes	Lesion localization and biopsy coordinates (including by type/sub-type of lesion) Tissue/fluid sampling result Needle path planning Reduced hematoma/bruising and bleeding Reduction in procedure time and reduction in multiple attempts to obtain tissue samples Radiation dose
Study types	Systematic reviews, randomised controlled trials, observational studies

2.2.2. Question about the incremental costs associated with implementing DBT

Research question 3 was:

What are the incremental costs associated with implementing DBT in the breast cancer assessment pathway compared to digital mammography?

The PICO(T/S) criteria underpinning research question 3 are described in *Table 4* (below).

Table 4: PICO(T/S) criteria for questions relating to the implementation of DBT

Criterion	Description
Population	Women aged over 40 years who undergo assessment for breast cancer after being recalled to assessment because of a suspicious area in the breast detected through a screening mammogram or women with symptoms <i>NB women with more dense/non-fatty breasts are a key sub-population group</i>
Intervention	DBT (either alone or when combined with s2DM)
Comparators	Digital mammography (including spot-compression or magnification views) Hand-held ultrasound
Outcomes	Work flow including image acquisition time and interpretation/reading time Technologist and radiologist training IT changes (including software/hardware upgrades and data storage)
Study types	Systematic reviews, RCT, observational studies, HTA, grey literature

2.3. Literature search

The following databases were searched in June 2018:

- Clinicaltrials.gov
- Cochrane Library database
- Embase
- National Institute for Health and Clinical Excellence
- OVID Medline
- ProQuest
- Scopus
- UK National Institute for Health Research HTA database, and
- UK NHSBPS.

To complete a systematic search, we used combinations of subject/index terms where appropriate (eg, exploded term ‘mammography’ or exploded ‘breast neoplasm’) in combination with key words, or key words alone depending on the search functionality of each database or website (eg, main searches included ‘tomosynthesis’ PLUS ‘breast cancer’ PLUS ‘assess*’ OR ‘diagnos*’ in the title or abstract).

The following limits were applied on all searches:

- a date criterion (1 January 2010 – 31 May 2018 or 2010 onwards)
- full English language manuscripts/articles
- human (i.e., studies using breast phantoms or specimen results were excluded), and
- study type restrictions (where available and appropriate, we restricted returns from research databases to peer-reviewed systematic reviews, literature reviews, RCT, observational studies and clinical trials).

Duplicate citations and a small number of false hits/inaccurate returns were removed before all initial returned citations and abstracts were reviewed for relevance to the main research questions. Material was excluded if it:

- did not relate to the role of DBT in the diagnosis or assessment of breast cancer (i.e., if it related to DBT as a population screening tool for breast cancer and did not contain information relevant to cancer characteristics)
- was covered in *Allen + Clarke's* literature review on the role of DBT in breast cancer screening
- focused on technical improvement to DBT (such the development of new reconstruction algorithms, new angles, resolution improvement, CAdE/search patterning, s-ray scattering, etc.), or
- related to second-look DBT if the lesion was MRI-only detected, related to intra-operative use of DBT, or follow-up surveillance post treatment.

To determine if this first search retrieved the correct range of available research, a validation process was completed using four recent systematic or literature reviews relevant to the primary research questions (Michell & Batohi, 2018; Destounis, 2017; Mall et al., 2017; Yun et al., 2017). There was a high degree of consistency between the studies returned using our strategies and those included in the four reviews. We are confident that the returned items include the main studies and articles that form the evidence base for the use of DBT in the assessment and diagnosis of breast cancer.

From this first sweep, full texts for all proposed inclusions were retrieved and reviewed for relevance to the research questions, inclusion criteria and documented PICOT criteria. A critical appraisal of study design (to determine overall quality) was completed and the bibliography of each included article was reviewed to identify other relevant research that may be of interest.

The citation review process for academic articles relating to the research questions is described in Figure 1 (overleaf).

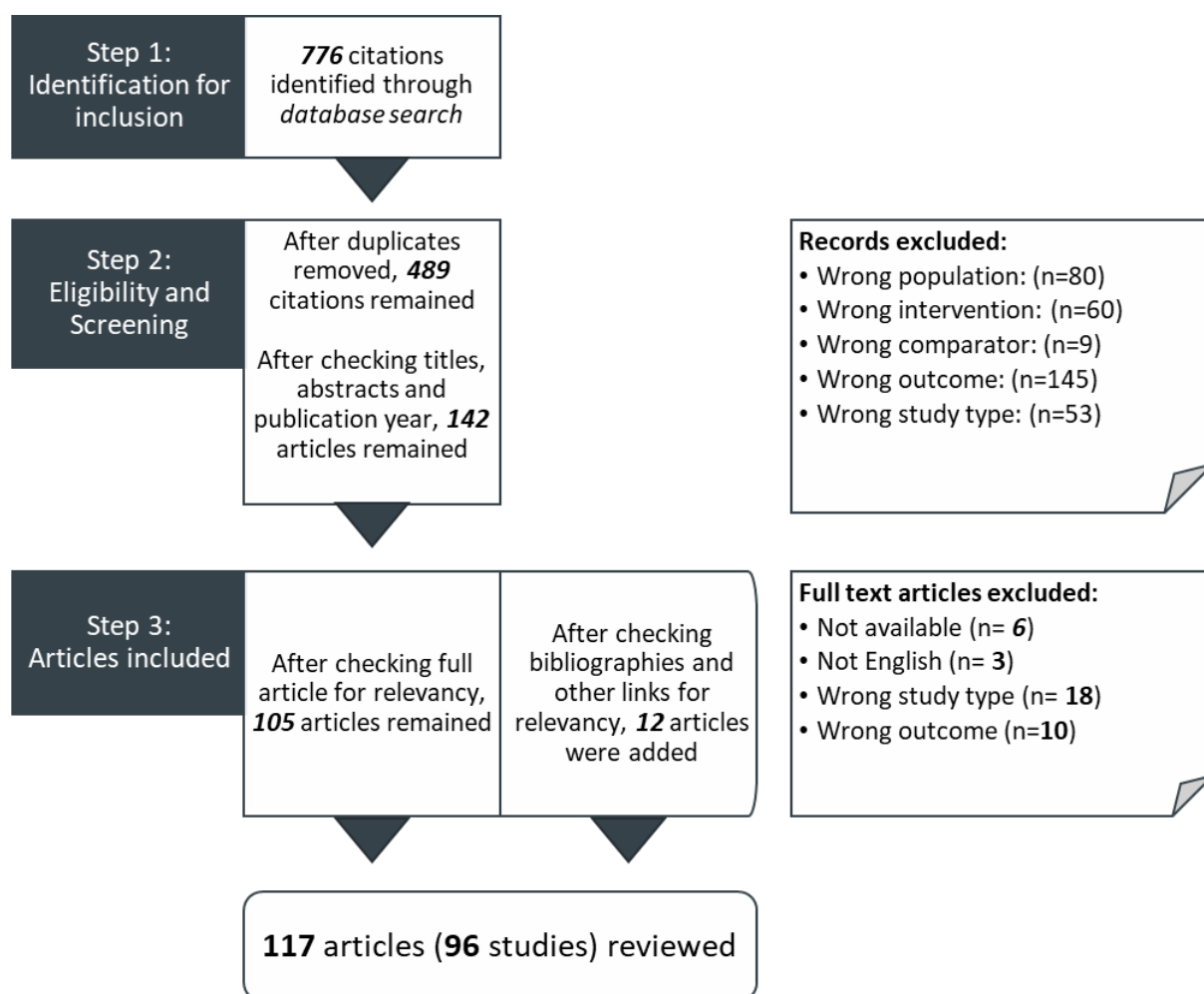


Figure 1: Citations review process

Study types were:

- four systematic reviews
- 14 narrative literature reviews
- no randomised controlled trials
- 96 articles (20 prospective studies; 76 retrospective studies)
- two position statements
- one technical or practical evaluation of DBT systems.

2.4. Limitations and interpretation

Developing an overall statement about the role of DBT is influenced by the following factors.

2.4.1. Limited Australian evidence creates uncertainty about the application of these research findings to the BSA program

Limited Australian evidence was identified in this review: only one oral abstract for Australian-based research (the Maroondah trial; Lockie et al., 2014) met the inclusion criteria for this study. Full publication of the results from the Maroondah study are expected to be published in the future. Further, DBT is currently used in some, but not all, assessment centres in the BSA

program. This creates uncertainty about applying the findings discussed in this literature review to the Australian context.

2.4.2. Lack of clarity about the imaging protocol used

Most studies described in this literature review compare DBT to FFDM alone or compare DBT to additional mammographic views such as DSCV. DBT modalities include either DBT alone or FFDM + DBT or DBT + s2DM compared to FFDM alone. However, many studies provide limited information about the actual imaging or reading protocol used. For example, authors:

- report the comparator modality as “DM” with no further description provided, and/or
- note that radiologists determine which mix of imaging techniques is needed, and/or
- do not provide further information in the method section that is sufficient to enable a clear understanding of the imaging protocol used or how it was used for specific suspicious mammographic presentations).

We have tried to provide as much detail about the DM imaging protocol as possible but recognise that in some cases “DM” was the only information provided. More limited information comparing the role of DBT to ultrasound or MRI in the assessment and diagnosis of breast cancer was returned in the searches. As such, this literature review presents information by main outcome (eg, diagnostic accuracy) with key differences in modality described where appropriate.

2.4.3. Study populations are mixed

Studies included in this literature review draw on different population groups:

- 20 studies only included women recalled to assessment for further work-up
- 15 studies included both women recalled to assessment and those with symptomatic presentation
- Nine studies included screening participants and women undergoing work-up for a symptom of breast cancer
- 21 studies only included women undergoing work-up for a symptom of breast cancer, or women who have already been diagnosed with breast cancer (mostly, these studies discuss tumour characteristics), and
- 19 studies provided insufficient detail about the clinical pathway that study participants have followed.

A breakdown of study by population is provided in Appendix A. The mixed study populations provide important information about the role of DBT given its use in working-up a suspected abnormality, regardless of how that concern arose (either through screening or symptomatic presentation) and may be necessary for adequate powering for rarer mammographic presentations or rarer breast cancer characteristics; however, they also make it more challenging to report on results that would be most relevant to a screening program. Inclusion of cancer-enriched cohorts may underestimate the difference in sensitivity as included cases may already have been detected with FFDM. In addition, cancer-enriched cohorts may limit the overall generalisability of findings to populations of women recalled to assessment.

In addition, most of the studies are retrospective observer studies with small sample sizes, many of which included either cancer-enriched samples or which had a mixed sample (i.e., women who were recalled to assessment, and those who had presented symptomatically).

Limited information about the powering required to detect significant results was usually described (i.e., only a handful of studies provided this information). Some studies provide a clear description of the exclusion criteria, recruited consecutive imaging work-up or have attempted to address potential bias; other studies are less clear, which makes it difficult to determine the overall influence of uncontrolled bias (where/if it exists). This may potentially influence the sensitivity/specificity findings that are not as reflective of a clinical environment where the population of women being assessed are only those recalled to assessment. It may also overstate the performance metrics for DM (as the lesion may have been first identified in DM at screening in studies using populations based on women recalled to assessment). Despite this, consistent findings from two meta-analyses provide some confidence in the overall direction and magnitude of effect in terms of diagnostic accuracy.

2.4.4. Reporting of mammographic results: use of BIRADS

The BSA program reports reading outcome using the National Breast Cancer Centre Synoptic Breast Imaging Report; however, most studies included in this literature review record reading outcome using the BIRADS system or the Royal College of Radiologists system for lesion classification to describe the accuracy of lesion classification (and, by extension, avoidable work-up for lesions with a benign final outcome) between different mammographic modalities. No research reported on the NBCC Synoptic Breast Imaging Report; however, the accuracy of reading outcome (using BIRADS or the College) is an important dimension of effectiveness of imaging technique and so we have included these study findings. Care is needed when considering the applicability of these findings to the Australian context.

2.4.5. Optimisation of DBT as a technique

Our initial search returned a large number (N=776) of papers relating to the optimisation of DBT as a technique (including optimisation of geometrical features, the role of CADe techniques, angle of image acquisition, reconstruction algorithms or other technical factors). Papers on second-look DBT were also returned. These studies were excluded from this literature review because, while meeting some of the PICOT criteria, they did not directly respond to the research questions or reported findings from studies that used breast phantoms and simulated images, which may affect the real-world application of the technology.

2.4.6. Inclusion dates

The current BSA position statement states that DBT provides benefit in the breast assessment clinic. *Allen + Clarke's* literature review had a date inclusion range of 1 January 2010 to current (June 2018). For some outcomes, this literature review reports on research published more recently than 2012 only because of the strength of the position taken in the existing BSA position statement in relation to the role of DBT in the assessment of breast cancer and given how much the evidence base has developed over the past five years. The full date inclusion range is used for other outcomes including area under the receiver-operating characteristic curve (AUC measurement), DBT's overall sensitivity and specificity compared to other views, its role in biopsy, and for other radiologic characteristics.

2.4.7. Terminology

The term 'reader' has been used to describe observer studies. Most readers are radiologists, but other readers are described when used (for example, the TOMMY trial included radiologists, advanced radiographers and breast physicians).

3. CAN DBT REPLACE DIGITAL MAMMOGRAPHY VIEWS IN THE BREAST CANCER ASSESSMENT CENTRE?

Women undergo assessment for breast cancer if they are recalled to assessment following the detection of an area suspicious for breast cancer on FFDM, or if they have symptoms suggestive of breast cancer. Accurate and timely assessment is critical to confirm a suspected breast cancer (or dismiss suspicious results as benign findings). Following confirmation, diagnosis is made based on the lesion type, lymph node status, biomarker, lesion size and metastasis. In general, assessment of abnormalities should include the least amount of imaging required (and therefore lowest radiation dose) to reach a conclusion. Swift confirmation or dismissal of malignancy can ensure timely, appropriate and effective treatment and can improve women's overall experience and prognosis/survival. *Figure 2* (below), reproduced from Mall et al. (2017), describes a general clinical pathway for breast cancer assessment and diagnosis.

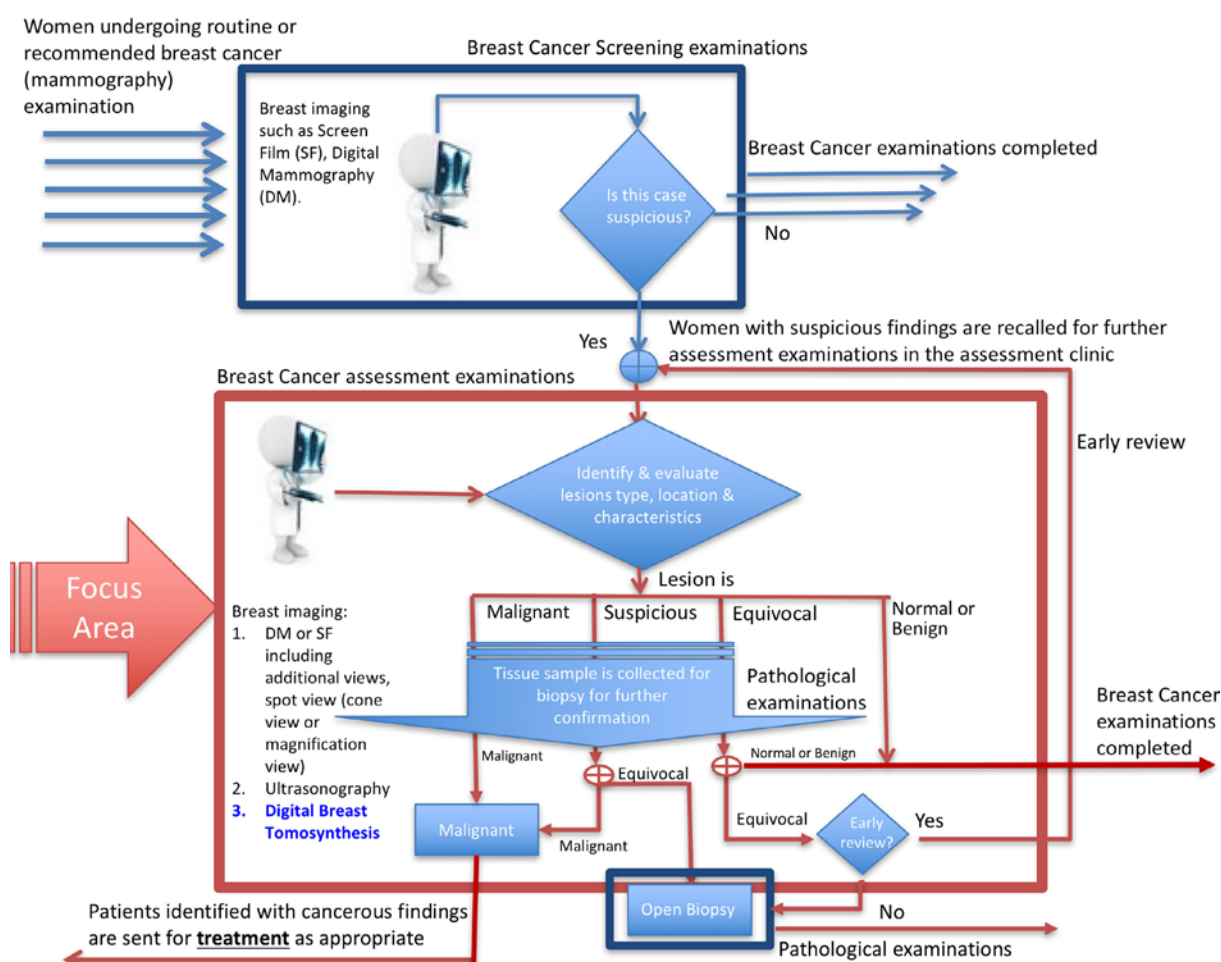


Figure 2: Breast cancer assessment pathway in a screening environment (reproduced from Mall et al., 2017)

In Australia, DBT is used in some BSA assessment centres (but not all). Where DBT is used, different BSA state and territory programs use it in a range of ways. For example, some programs have clinical guidelines requiring the use of DBT in the work-up of all non-calcified lesions; others require the radiologist to determine when DBT should be used. Some states do not offer DBT in all their assessment centres. Others do not offer DBT at all.

We want to know, based on current evidence, what role DBT can play in the assessment and diagnosis of breast cancer in women aged over 40 years. We want to know, based on current

evidence, if DBT is a more sensitive, specific and safer imaging technology for the work-up of screen-detected abnormalities (i.e., its potential role in the assessment and diagnosis of breast cancer within a breast screening program) compared to digital mammographic views, including repeat FFDM, DSCV, or adjunctive ultrasound or adjunctive MRI. Chapter 3 describes the current evidence base comparing DBT's performance compared to these other imaging work-up views. This information is used to answer Research Question 1 about the effectiveness and safety of DBT in the assessment clinic and provide a statement about the quality of evidence underpinning the answers to these questions. We describe the evidence on:

- diagnostic accuracy of DBT compared to DM (usually presented as an area under the ROC curve value based on sensitivity and specificity findings)
- inter-observer agreement
- accuracy of overall mammographic reporting results (i.e., Breast Imaging Reporting and Data System (BIRADS) classification agreement), and
- radiation dose.

3.1. Diagnostic accuracy of DBT imaging compared to repeat FFDM or other DM imaging (clinical pathway not specified)

Sensitivity: the proportion of breast cancers correctly identified by a diagnostic test or the true positive rate.

Specificity: the proportion of women correctly identified as not having breast cancer or the true negative rate.

Increased sensitivity with DBT (compared to DM) indicates that DBT correctly identifies more breast cancers compared to other imaging modalities. Increased specificity would indicate that DBT results in the correct dismissal of more women with normal or benign findings without the need for further work-up (i.e., a reduction in false positive recalls for further assessment). A test's diagnostic excellence can be measured by the area under ROC curve or JAFROC analysis (with results presented as AUC measurement). AUC measurement summarises the trade-off between sensitivity and specificity and represents how accurately a test assesses a person's probability of having cancer. An excellent diagnostic test has an AUC measurement of between 0.9 and 1. Good tests report AUC measurement of between 0.8 and 0.9.

Section 3.1 discusses literature comparing diagnostic accuracy of imaging protocols involving DBT compared to DM (i.e., repeat FFDM, other DM views or DSCV), adjunctive ultrasound or adjunctive MRI. Two sets of studies are described:

- DBT compared to repeat FFDM or DM (not further specified, that is DM is the only descriptor provided and the imaging protocol is not clear from the study description), and
- DBT compared to digital spot compression views (DSCV).

The evidence base is complicated by researchers attempting to determine the optimal way to use DBT in a way that balances improved accuracy with a safe (and lowest) radiation dose. This means different ways of using DBT are reported:

- FFDM + DBT compared to FFDM or DM (not further described), and
- DBT alone or in one view only compared to FFDM or DM (not further described).

Most of the evidence is based on small retrospective studies, including those using cancer-enriched datasets and different DBT-capable units (including prototypes). While the overall direction of effect leans toward increased diagnostic accuracy with DBT, study design limitations may affect applicability to clinical settings within a breast cancer assessment centre. Further research is needed to remove the impact of bias inherent in retrospective observer studies and to determine the optimal way of using DBT. As research seeks to determine the optimal way of implementing DBT in assessment and diagnosis, studies reported on diagnostic accuracy for a range of imaging protocols including:

- FFDM + DBT compared to FFDM
- FFDM + DBT_{MLO} compared to FFDM
- different 1v-DBT combinations compared to DM (unspecified or one view), or
- s2DM compared to FFDM.

That said, there is sufficient evidence of improved diagnostic accuracy and sensitivity for some screening programs (including the NHS BSP, Ireland's screening program, and some state BSA programs) and private clinics to have already moved to using DBT in the assessment of lesions suspicious for breast cancer.

The three systematic reviews, one narrative literature review and 24 studies discussed in this literature review are listed below. Most of these studies included mixed study populations (i.e., women recalled to assessment and symptomatic women or screening and symptomatic women or symptomatic women only). Summaries of study results are included in *Table 5*, *Table 6* and *Table 7*. Some small studies with significant methodological limitations are included in these tables but are not described in the narrative in this section. In addition, primary studies discussed in the systematic reviews are not further described unless additional material from the primary study was more useful when described in full.^{8,9}

Systematic reviews and meta-analysis

Two systematic reviews with meta-analysis: Phi et al. (2018); Lei et al. (2014)

One narrative systematic review: Garcia-León et al. (2015)

One narrative literature review: Michell & Batohi (2018)

RCTs and prospective studies

RCTs: none

Four prospective studies:

Women presenting for screening / symptomatic women: Mansour et al. (2014)

Only symptomatic women: Cai et al. (2015); Mercier et al. (2015)

Diagnostic pathway unknown: Alakhras et al. (2014)

⁸ This literature review includes additional data on diagnostic accuracy from the following studies included in Garcia-León et al.'s systematic review (without meta-analysis): Brandt et al. (2013); Rafferty et al. (2013); Waldherr et al. (2013); Noroozian et al. (2012); Tagliafico et al. (2012); and Wallis et al. (2012). Data from the TOMMY trial (Gilbert et al., 2015a,b) was included in Phi et al.'s meta-analysis; however, we have also described findings from this study in our literature review given that it is the largest and one of the most significant studies reporting on diagnostic accuracy in population-based screening program.

⁹ Thirteen of the 16 studies included in Phi et al.'s systematic review were discussed in *Allen + Clarke's* literature review on the role of DBT in breast cancer screening. The remaining three studies are discussed in greater detail in this literature review (Chae et al., 2016; Gilbert et al., 2015a; Waldherr et al., 2013).

Retrospective studies

20 retrospective studies:

Women recalled to assessment: Choi et al. (2016a); Gilbert et al. (2015a); Haq et al. (2015)

Women recalled to assessment and symptomatic women: Garayoa et al. (2018); Kang et al. (2016); Seo et al. (2016); Rafferty et al. (2013); Waldherr et al. (2013)

Women presenting for screening / symptomatic women: Mansour et al. (2014); Rafferty et al. (2014)

Only symptomatic women: Tang et al. (2017); Bian et al. (2017); Mariscotti et al. (2016); Taha Ali et al. (2016); Bansal & Young (2015)

Diagnostic pathway unknown: Endo et al. (2018); Chan et al. (2017); Mariscotti et al. (2017); Thomassin-Naggara et al. (2015); Gennaro et al. (2013a)

Key findings

This literature included three systematic reviews (covering 34 studies), one narrative literature review and 24 primary studies comparing the assessment performance of DBT to DM views. DBT's assessment performance and diagnostic accuracy has not been robustly assessed in RCT or large prospective studies in clinical settings. It is also unclear if all available studies have sufficient powering to detect changes in diagnostic performance.

Both DBT and DM are excellent diagnostic tests with DBT appearing to have diagnostic superiority (as measured by AUC) but this finding is not definitive

There is a large body of evidence demonstrating benefit when DBT is used to assess findings suspicious for breast cancer (regardless of imaging modality, study design, sample or DBT unit). Both DM and DBT had very good to excellent overall diagnostic accuracy (based on a measurement of area under the receiver operating characteristic curve). Using ROC or JAFROC analysis to assess how accurately DBT and DM assess a woman's probability of having cancer, evidence (including pooled analysis from three systematic reviews), demonstrated a consistent (if varying) increase in DBT's performance. Increase in AUC measurement from combinations of/comparisons between DBT and DM from pooled analysis were reported at an AUC measurement of 0.8668 for one-view DBT compared to 0.8561 for DM. AUC measurement observed for FFDM + DBT ranged from 0.788 to 0.914 compared to 0.681 to 0.881 with FFDM. Median AUC measurement fell between 0.8-0.9 for both modalities, including results presented from the TOMMY trial with overall AUC measurement for DBT imaging protocols more likely to exceed 0.85. Overall, FFDM + DBT appears to provide at least equivalent assessment performance and diagnostic accuracy compared to FFDM, with some studies indicating superior (but not perfect) performance. Diagnostic performance with comparing a reconstructed 2D image (s2DM) shows promising performance results compared to FFDM.

Some screening programs and clinics have already implemented DBT into assessment

While there are limitations in the evidence base (small sample sizes, retrospective designs, cancer-enriched samples, use of DBT-capable prototypes, etc.), findings presented in the primary studies have been sufficient to drive changes in clinical and screening program practices (both internationally and within the BSA program) in relation to the assessment of lesions suspicious for breast cancer. DBT's superiority in terms of AUC measurement and sensitivity is often cited as the reason for change.

3.1.1. Some screening programs and clinics have already implemented DBT into the assessment pathway

It is important to note that reported diagnostic equivalence or superiority for DBT compared to DSCV has already resulted in practice changes and DBT now replaces DSCV in some screening programs including the BSA and the NHSBSP (in locations where DBT-capable units are available) (Ni Mhuircheartaigh et al., 2017) as well as in individual clinical practices (for example, Peppard et al., 2015). While Bansal & Young (2015) reported a non-significant improvement in AUC measurement, the reported improvement resulted in the replacement of DSCV with DBT in the work-up of non-calcified lesions in Bansal & Young's clinical practice (even though they did not compare DBT to DSCV).

3.1.2. Both DBT and DM are excellent diagnostic tests with DBT appearing to have diagnostic superiority (as measured by AUC) but this finding is not definitive

Systematic reviews

Three systematic reviews reported on diagnostic accuracy for DBT compared to 2D mammography. Two systematic reviews included meta-analysis.

Two of the systematic reviews (Garcia-Léon et al., 2015; Lei et al., 2014) were based on early studies including studies set in a primary screening environment as well as those enrolling asymptomatic/symptomatic study participants from breast cancer assessment clinics. Studies were completed using both commercially available DBT-capable units and prototype units. These systematic reviews generally included few large, prospective studies from assessment settings and which compared diagnostic accuracy of DBT and mammography (most of the large prospective trials are set in a primary screening setting). Study design limitations in the primary literature led to significant heterogeneity and limited the extent to which pooled analysis could be completed.

Garcia-León et al. (2015) completed a systematic review to evaluate the effectiveness and diagnostic validity of DM compared to DBT. The authors included 11 studies (published between 2010 and 2013). Studies were undertaken in either a screening or diagnostic setting. Women were worked-up with DBT (either Hologic's Selenia Dimensions or prototype units) and compared to either FFDM, compression and magnification views, or ultrasound; however, no further specific details about the imaging pathway used in each primary study in a diagnostic setting were provided. Primary No pooled analysis was completed due to the heterogeneity of included studies. The authors also noted that the included studies had a high risk of bias and likely overstated the overall diagnostic accuracy of DBT because of the use of cancer-enriched populations: some studies only included participants with lesions with suspicious for malignancy (i.e., BIRADS classifications of 4 and above), some excluded women with mammographic features like calcifications, etc. Results from the studies set in diagnostic populations are described below.

Lei et al. (2014) is another older systematic review with meta-analysis that assessed DBT's diagnostic performance. The main inclusion criteria in Lei et al.'s systematic review were studies published between 1950 and June 2013 using DBT and DM for diagnosis of lesions suspicious for breast cancer (BIRADS reading report of 3 or higher). Primary studies had to have a sample size of more than 30 participants with biopsy-confirmed histological results and adequate data from which to calculate sensitivity and specificity measures. The authors included five retrospective and two prospective European or American studies (2104 women) and noted that while studies were high-quality (using QUADAS assessment), there was significant

heterogeneity. Most of these studies were undertaken in a diagnostic setting, and these results are described below.

Because of the increased radiolucency of dense breast tissue and overlapping of tissue structures, the overall sensitivity and specificity of mammography is reduced in women with more dense breasts (BIRADS density classification 3 or 4/c or d). Phi et al.'s 2018 systematic review with meta-analysis looked at diagnostic parameters of DBT compared to DM in women with more dense breasts. The main inclusion criteria were studies with women aged 18 years or older who had dense breasts and who had undergone breast imaging with DBT and DM. A range of imaging combinations were used in the primary studies (i.e., 1v-DBT compared to FFDM; DBT compared to FFDM; 1v-DBT + FFDM compared to FFDM). Phi et al. (2018) reported on 16 studies published between May 2007 and May 2017, 11 of which were set in a primary screening environment (including some of the key papers from *Allen + Clarke's* literature on the role of DBT in screening). Five studies were from diagnostic settings and Phi et al. (2018) analysed these results separately.

There was limited cross-over in terms of the studies selected for each systematic review: from a total of 34 studies, only six were reported in more than one systematic review. Primary studies incorporated into the systematic reviews were read as part of this literature review and, in some cases, findings are separately discussed where further information was available that was not separately or clearly described in systematic review. Results from diagnostic studies in each systematic review are presented in *Table 5*.

Systematic review and meta-analysis results: diagnostic accuracy

FFDM + DBT compared to FFDM alone

Based on findings published up to 2013, Garcia-León et al. (2015) and Lei et al. (2014) reported that DBT is an excellent diagnostic test (as were the mammography-based tests); however, whether DBT alone is a superior diagnostic test compared to DM remains uncertain. This finding is based on studies published before 2013. Looking at data from studies from diagnostic populations only, Garcia-León et al. reported that DBT was an excellent test. Its greatest utility was adjunctive (i.e., FFDM + DBT):

- Reported AUC measurement was highest in women whose abnormality was detected at screening and when FFDM + DBT was used (AUC measurement: 0.96; 95% CI: 0.95, 0.97), and
- Lowest reported AUC measurement was when the woman had a symptomatic presentation and was assessed using DBT alone (one or two views) (DBT alone: AUC measurement: 0.91), but
- Data on DM AUC measurement was not presented by the authors, making it challenging to compare performance.

Garcia-León et al. also reported on negative OR, noting that DBT was good or excellent (no data provided). The authors concluded that, based on these early studies, DBT could be useful as a diagnostic confirmation test that would reduce the use of other mammographic work-up views, but the authors considered that DBT's overall superiority compared to DM was unclear due to the low quality of the primary studies.

Lei et al. reported pooled summary ROC AUC measurement that favoured 1v-DBT (not specified) over DM (0.8668 compared to 0.8561). From the diagnostic studies, they also reported pooled

Q* indices¹⁰ of 0.7973 (DBT) and 0.7870 (DM). Lei et al. concluded that as DBT provides better visualisation of lesions, it provided improved discrimination between breast cancers and benign/normal findings compared to DM. The diagnostic OR for DBT was 26.04 compared to 16.24 for DM ($p=.0000$) and pooled positive likelihood ratio was higher, suggesting that DBT has better discriminatory ability for breast cancer.

Underpinning the AUC measurement is data on sensitivity and specificity. The three systematic reviews reported that DBT increased sensitivity when compared to DM:

- Lei et al.'s analysis discussed all results as relating to 1v-DBT but it is not clear whether the underpinning studies all investigated this particular use of DBT or if the population in which women were drawn from was diagnostic only (although at least some, including Thibault et al.'s 2013 study, reported on DBT_{MLO} for women recalled to assessment and those presenting with symptoms). The impact of this uncertainty makes it difficult to determine the overall sensitivity of DBT when used diagnostically. Pooled analysis of seven studies using histological results as the reference standard showed a pooled sensitivity of 90% (95% CI: 87, 92) for DBT compared to 89% (95% CI: 86, 91).
- Primary studies set in diagnostic populations reported by Garcia-Léon et al. all reported an increase in sensitivity with reported sensitivity in the primary studies ranging from 69% to 100%. 95% CI ranges were provided for each included study, but it is not possible to tell the DM values as these were not presented. No pooled analysis was completed.
- Phi et al. (2018) reported that, in diagnostic populations, DBT had higher sensitivity compared to DM in diagnostic settings for women with dense breasts: 84-89% for DBT compared to 69-86% for DM images.

Both Garcia-Léon et al. and Lei et al. reported increased specificity with DBT compared to DM. In Phi et al.'s study, reported specificity rates were similar between DBT and DM imaging for women with more dense breasts (ranging from 72-93% with DBT alone or when used as an adjunct to DM, compared to 57-94% for DM alone).

While issues with different imaging combinations, selection of population in the included primary studies, test timing and recall protocols are reported, studies set in diagnostic populations included in Phi et al.'s 2018 analysis reported a risk ratio from a diagnostic study favouring DBT over DM (RR 1.12; 95% CI: 1.01, 1.24). Phi et al. concluded that DBT (whether alone or with DM) increased sensitivity but not specificity.

Table 5: Results from systematic reviews assessing AUC measurement, sensitivity and specificity

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Phi et al. (2018)	16 prospective and retrospective studies published to May 2017 and which investigated DBT and DM in a screening (asymptomatic) (11 studies) or diagnostic/recalled to assessment setting (5 studies) Participants: 116,062 women who had DM breast imaging;	ROC plot data provided indicating consistently higher values for DBT compared to FFDM. No explanation or specific figures provided.	<i>Reported ranges in primary studies from diagnostic settings</i> FFDM: 69-86% (58, 89) DBT: 84-89% (71, 95)	<i>Reported ranges in primary studies from diagnostic settings</i> FFDM: 57-94% (55, 97) DBT: 72-93% (68, 96)

¹⁰ The optimal point on the ROC curve.

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
	55703 women who had DBT imaging Cancers: DM = 1627 malignancies; DBT = 1482 malignancies Unit: not described			
Garcia-León et al. (2015)	11 prospective studies published between June 2010 and Feb 2013 investigating screening and diagnostic validity of DBT and FFDM. 10 were cohort studies and 1 was a case-control study. 3 studies were set in clinical settings only, 4 were set in clinical and screening settings; and 4 were set in screening only. Participants: 2475 women with either abnormalities detected at screening or clinical signs of breast cancer. Cancers: not stated Unit: Hologic Selenia Dimensions used in 6 studies; prototype units used in 5 studies	<i>AUC for women with clinical symptoms</i> DBT (1 or 2 views): 0.91 1v-DBT: 0.86 (range: 0.80-0.91) <i>Women at screening</i> FFDM + DBT: 0.96 (range: 0.95-0.97) <i>Women with dense breasts</i> DBT: 0.85 DBT (1 view): 0.77 <i>NB p-values or 95% CI not provided for most results</i>	<i>DBT sensitivity data from diagnostic setting studies only</i> 69.8-89.7% <i>NB sensitivity data by intervention and comparator not provided</i>	<i>DBT specificity data from diagnostic setting studies only</i> 54.4-88.9% <i>NB specificity data by intervention and comparator not provided</i>
Lei et al. (2014)	Five prospective and two retrospective studies Participants: 2014 women with 2666 lesions classified as ≥ 2 BIRADS Cancers: 640 TP/FN Unit: not described but presumably a range of units were used	<i>Pooled SROC AUC measurement</i> DM: 0.8561 1v-DBT: 0.8668 <i>Pooled SE (AUC) measurement</i> DM: 0.0718 1v-DBT: 0.0573	<i>Pooled sensitivity</i> DM: 89% (86, 91) 1v-DBT: 90% (87, 92) <i>Diagnostic OR</i> DM: 16.24 (5.61, 47.04; $p=.0000$) 1v-DBT: 26.04 (8.7, 77.95; $p=.0000$) <i>Pooled positive likelihood ratio</i> DM: 2.82 (1.77, 4.52) 1v-DBT: 3.59 (2.31, 5.30)	<i>Pooled specificity</i> DM: 72% (70, 74) 1v-DBT: 79% (77, 81) <i>Pooled negative likelihood ratio</i> DM: 18% (9, 38) 1v-DBT: 15% (6, 36)

The systematic reviews established DBT's diagnostic equivalence compared to 2D mammography in terms of AUC measurement and sensitivity, with some mixed results for specificity. The remainder of this section discusses whether later studies (i.e., published from 2014 onwards) from diagnostic settings also demonstrated equivalent or superior diagnostic accuracy, unless it is reporting additional findings from studies included in any of the systematic reviews.^{8,9}

FFDM + DBT compared to FFDM: diagnostic accuracy

Five prospective studies published since 2013 reported on diagnostic accuracy for FFDM compared to FFDM + DBT. None of these studies included only asymptomatic women recalled to assessment.

Prospective studies from mixed populations

Alakhras et al. (2014) provided further evidence of FFDM + DBT's improved diagnostic accuracy (expressed as increase in AUC measurement and increased sensitivity) compared to FFDM.

Mixed results were presented for specificity. Using a test set (50 images, 32 malignancies), with FFDM images interpreted first, then FFDM + DBT images, AUC results were 0.788 for FFDM + DBT compared to 0.681 for FFDM (Wilcoxon test statistic -337.0, $p < .0001$). The authors also reported on JAFROC figure of merit (FOM) analysis, finding that FFDM + DBT had a statistically significant higher JAFROC FOM compared to FFDM (0.745 compared to 0.621, Wilcoxon test statistic -274.0, $p < .0001$). In this study, FFDM + DBT also increased sensitivity significantly (from 63% with FFDM to 70.4% with FFDM + DBT) and specificity (from 65.2% with FFDM to 78.3% with FFDM + DBT).

While not reporting diagnostic accuracy using AUC measurement, Mercier et al. (2015) used a cancer-enriched sample to consider sensitivity and specificity in a prospective study comparing DBT to DM, ultrasound and MRI. The authors concluded sensitivity and PPV increased with the use of DBT compared to DM, but specificity was lower for DBT with equal NPV (i.e., there was a higher rate of false positive classifications with DBT 74.4% compared to 81.6%). Lower (but not significant) specificity with DBT may have been due to reader bias as readers were aware of clinical presentation and staging of a BIRADS 4 or 5 lesion (and therefore the likelihood of cancer) and were therefore more likely to positively grade DM images.

None of Tang et al.'s (2017) results achieved statistical significance but in this study of symptomatic women, 197 consecutive women (153 malignancies) had imaging work-up with FFDM compared to FFDM + DBT. The AUC measurement was higher overall for FFDM + DBT compared to FFDM (over 0.9 and 0.8 respectively). Tang et al. also reported higher sensitivity with FFDM + DBT compared to FFDM (71.8% and 71.2% with FFDM for reader 1 and reader 2 compared to 92.1% and 90.8% with FFDM + DBT). PPV and NPV were also higher with FFDM + DBT compared to FFDM. Specificity results were mixed: it was lower for FFDM + DBT compared to FFDM for one reader but higher for the other reader.

Two studies (Cai et al., 2015; Mansour et al., 2014) had significant methodological limitations associated with small sample sizes, included both screening and diagnostic populations, and limited description of results or issues with the selected study cohort: results are presented in *Table 6* only. Results from Tang et al.'s study are included in *Table 18*.

Retrospective studies

Eight retrospective studies reported on the diagnostic accuracy of FFDM + DBT compared to FFDM (six using Hologic Selenia Dimensions and two used a Hologic prototype as the data was collected before FDA approval). In seven studies, diagnostic accuracy significantly increased when FFDM + DBT was used. The other study reported a non-significant increase.

In studies achieving a significant result, the AUC measurement range with FFDM + DBT was 0.827 to 0.901; sensitivity ranged from 76.2% to 91.9%. For FFDM, the AUC measurement range was 0.721 to 0.84; sensitivity ranged from 61% to 87%. Mixed overall results were presented for specificity, with four studies reporting (generally non-significant) increases of between 0.2 to 13 percentage points, equivalence (one study) and three studies describing small decreases (1.4 to 2 percentage points). Increases in PPV and NPV with FFDM + DBT were reported in two studies. Together, this suggests that FFDM + DBT results in more accurate identification of women with breast cancer and increased correct dismissal of normal or benign imaging. Individual study results are described below.

Retrospective studies of women recalled to assessment

The largest study discussed in this literature review is the TOMMY trial. The TOMMY trial involved 26 radiologists, advanced radiographers and breast physicians who retrospectively reviewed a cancer-enriched dataset drawn from a multicentre trial of 6020 women recalled to

assessment for assessment in the NHSBSP and 1040 women aged under 50 years attending annual screening because of family history of breast cancer (Gilbert et al., 2015a). There were three reading arms: FFDM, FFDM + DBT and DBT + s2DM. Overall diagnostic performance for both DBT arms (FFDM + DBT or DBT + s2DM) was slightly superior compared to FFDM. AUC measurement was 0.89 FFDM + DBT and 0.88 for DBT + s2DM, compared to 0.84 for FFDM ($p<.001$). Gilbert et al. reported that the greatest improvement in diagnostic performance was seen in women with more dense breasts (defined as $>50\%$, not further described): AUC measurement for women with more dense breasts were 0.83 (FFDM) compared with 0.89 (FFDM + DBT), and 0.87 (DBT + s2DM) (difference between DBT reading arms and FFDM, $p<.001$). It is interesting to note that the AUC measurement for FFDM + DBT for women with more dense breasts is higher than the overall FFDM AUC measurement for all women.

Sensitivity data from the TOMMY trial indicated that FFDM + DBT correctly detected more cancers than FFDM and recalled fewer benign lesions to assessment compared to FFDM. From a total of 1137 cancers, 921 cancers were detected with both FFDM and FFDM + DBT, 71 were detected only with FFDM, 95 were detected only with FFDM + DBT. A total of 50 were missed in either reading arm. No results were presented for DBT + s2DM nor is any information about cancer type detected in each reading arm presented. Overall, this suggests that cancer detection may be increased with FFDM + DBT, a large number of cancers ($n=121$) may still only be detected with FFDM and that FFDM is still an important imaging modality in cancer detection.

Reported statistical analysis from the TOMMY trial found that DBT conferred an increase in OR of 34% (1.34, 95% CI: 0.97, 1.85, $p=.06$) with a 56% reduction in recall of non-cancers (OR 0.44; 0.39, 0.49, $p=.001$).

Other findings were that:

- FFDM + DBT significantly improved sensitivity compared to FFDM (89% compared to 87%) and sensitivity was higher for women aged 50-59 years and for women with more dense breasts
- DBT significantly improved specificity (from 58% with FFDM to 71% with DBT + s2DM, $p=.001$) and this increase was observed in all density subgroups and by primary mammographic feature, which would lead to a 19% decrease in benign lesions considered to be malignant on FFDM imaging
- overall sensitivity for lesions with a primary mammographic presentation of microcalcifications was lower overall; however, there was higher specificity in the DBT reading arms: 31%, 39% and 44% ($n=745$ cases, 750 cases and 723 cases for FFDM, FFDM + DBT and DBT + s2DM respectively).

Based on this large cancer-enriched sample largely drawn from a screening population, the authors concluded that either FFDM + DBT or DBT + s2DM provided a clear diagnostic improvement compared to FFDM. In addition, DBT sensitivity may be underestimated as participants were selected due to abnormal FFDM screening findings (i.e., lesions were already visible on FFDM).

Retrospective studies of women recalled to assessment or symptomatic women

Similar significant results were reported by Seo et al (2016) but the reported increases in AUC measurement and sensitivity were larger than those reported in the TOMMY trial and were slightly higher overall. Seo et al. used JAFROC analysis to compare the likelihood of malignancy rating, reporting statistically significant improvements with FFDM + DBT and DBT alone compared to FFDM. Sensitivity with this modality also increased ($p<.001$) as did overall specificity (but significance was not achieved, and individual reader results were mixed, which

the authors attributed to the high proportion of cancer cases in the sample population). Seo et al (2016) concluded that FFDM + DBT offered superior diagnostic accuracy compared to FFDM alone.

In their first observer study, Rafferty et al. (2013) used a cancer-enriched dataset of 312 cases in which readers used a recall/non-recall ranking for each image (firstly blinded, then with clinical history and prior imaging). Pooled AUC measurement demonstrated that FFDM + DBT had a higher overall diagnostic accuracy compared to DBT alone: a statistically significant increase in AUC measurement of 7.2% overall and for non-calcified lesions of 10.4% (both $p < .001$). Sensitivity and specificity also increased but results were still lower than the TOMMY trial. Proposed reasons for the gain in diagnostic accuracy were that DBT allowed better imaging of multi-focality, reduced tissue superimposition and more accurate evaluation of mass margins, making it easier to accurately determine malignancy or otherwise. AUC measurement did not reach significance for calcified lesions. Although diagnostic accuracy was improved, Rafferty et al. also reported that one cancer (ILC) was missed on DBT but detected on FFDM. To improve reader performance, additional reader training was completed, and a second study undertaken.

Waldherr et al.'s (2013) study, where for women with BIRADS 3 or 4 breasts, sensitivity and specificity increased with DBT_{MLO} alone (84% and 83.9% respectively) compared to the reported sensitivity and specificity with FFDM in women with less dense breasts (BIRADS 1 or 2) (78.6% and 73.3% respectively).

Retrospective studies of women presenting for screening and symptomatic women

In Rafferty et al.'s second observer study (Rafferty et al., 2014), the authors compared:

- FFDM + DBT_{MLO} to FFDM, and
- FFDM + DBT to FFDM.

Using a cancer-enriched dataset of 310 women (51 malignancies) and a DBT prototype unit, radiologists first reviewed FFDM images then DBT_{MLO} then DBT. Improvements in AUC were reported for all readers when DBT or DBT_{MLO} was used. Overall, both DBT reading protocols had statistically significantly higher AUC measurement compared to FFDM, with increases in AUC measurement of 0.036 reported for FFDM + DBT_{MLO} and 0.068 for FFDM + DBT (the two-view combination performed better than the DBT one view combination). The authors also reported AUC measurement for women with more and less dense breasts (BIRADS 4th edition). AUC increased significantly with the use of DBT in either one or two views:

- AUC for non-dense breasts (FFDM; FFDM+DBT_{MLO}; FFDM+ DBT): 0.880, 0.898, 0.915, and
- AUC for dense breasts (FFDM; FFDM+DBT_{MLO}; FFDM+ DBT): 0.786, 0.832, 0.877.

This increase in AUC measurement with FFDM + DBT aligns with Gilbert et al.'s finding even though the clinical pathway differed slightly (that with FFDM + DBT, diagnostic accuracy for women with dense breasts is more comparable to the level of accuracy for women with less dense breasts when using FFDM).

In Rafferty et al.'s 2014 study, DBT (whether in one view or two) detected more cancers than FFDM (an increase in sensitivity of 8.7 percentage points with DBT_{MLO} and 16 percentage points for DBT was reported). In both Rafferty et al. studies, sensitivity increased the most for invasive cancers (study 2 = 21 percentage points with DBT; study 1 = 15%) compared to in situ cancers

Fewer non-cancer cases were recalled with both DBT and DBT_{MLO} compared to FFDM, with a statistically significant decrease seen with both DBT views (the greatest being with DBT rather

than DBT_{MLO} from 54.2% to 33.5% with FFDM). Also related to this was an increase in the number of cancer cases recalled with DBT compared to FFDM (90.2% to 84.3%). In combo mode (FFDM + DBT) a reduction in recalls was also reported. Mean recall/non-recall rates for benign lesions were 48.8% with FFDM compared to 30.1% with FFDM + DBT. Similar recall rates for malignancies were reported (84.8% compared to 85.7%). PPV and NPV both increased with the addition of DBT.

Retrospective studies of symptomatic women only

In Mariscotti et al.'s (2016) study, AUC measurements for FFDM + DBT compared to FFDM were reported in the detection of invasive lobular carcinoma (ILC, a cancer that can be difficult to detect with mammography and a cancer that is more likely to present as a subtle mammographic finding like asymmetry or architectural distortion, AD). While a small study (83 women with 107 ILC), it is important because it provided statistically significant evidence of FFDM + DBT's superior performance in detecting cancer with a subtle mammographic presentation. AUC measurement for FFDM + DBT were 0.89 compared to 0.84 with FFDM ($p < .0001$). In this study, sensitivity was also increased but specificity was slightly lower overall (but the cohort consisted only of ILC cases). More information about ILC detection is included in *Part 5*.

Bansal & Young (2015) reported a non-significant improvement in AUC measurement favouring FFDM + DBT compared to FFDM (0.901 compared to 0.721) in a non-cancer enriched sample of 103 symptomatic women with subtle, non-calcified mammographic presentations. Bansal & Young's study is important as it did not use an enriched cohort.

FFDM + DBT_{MLO} compared to FFDM: diagnostic accuracy

FFDM + DBT results in a higher radiation exposure for women compared to FFDM. As each DBT acquisition has a lower mean glandular dose (MGD) than FFDM, the hypothesis was that DBT alone or in one-view combination could result in lower radiation exposure to women undergoing assessment or diagnostic work-up without a drop in diagnostic accuracy.

Three studies (including Rafferty et al., 2014, described above) investigated a different adjunctive use of DBT: FFDM + DBT_{MLO} compared to FFDM. All studies reported results that are consistent with the results presented for FFDM + DBT compared to FFDM. That is, there was an increase in diagnostic accuracy with similar AUC measurement ranges presented even though the imaging protocols differed. The AUC measurement with FFDM + DBT_{MLO} ranged from 0.809 to 0.914. A small range of sensitivity values were presented compared to the FFDM + DBT results (ranging from 83 to 89.3%); specificity ranged from 50.7% to 94%. For FFDM, the AUC measurement range was 0.692 to 0.881; sensitivity was about 73% in all studies; specificity ranged from 50.5% to 90%.

Retrospective studies: symptomatic women only

Kang et al. (2016) (study described in the following sub-section) also reported on a reading arm of FFDM + DBT_{MLO} but did not report that this imaging protocol improved diagnostic accuracy. Overall, FFDM + DBT_{MLO} had the highest AUC measurement and sensitivity (0.914 and 83%) of all four protocols studied but there was no statistically significant difference between FFDM and FFDM + DBT_{MLO}. Specificity was also improved in relation to FFDM (see below for further information on the two other arms reported in Kang et al.'s study). Kang et al. did not report PPV or NPV.

Retrospective studies: diagnostic pathway unknown

Thomassin-Naggara et al. (2015) retrospectively evaluated readers' performance in assessing 155 FFDM and DBT_{MLO} images (61 breast cancers; 94 benign or high-risk lesions). The authors reported that FFDM + DBT_{MLO} resulted in a statistically significantly improved diagnostic performance compared to FFDM: AUC measurement was a little lower than other studies using a combination mode (from 0.692 with FFDM to 0.809 with FFDM + DBT_{MLO}). There was a corresponding increase in sensitivity, specificity and PPV/NPV. This increase in detection/probability of malignancy with FFDM + DBT_{MLO} led the authors to conclude that there was a decrease of 6.5% to 26.2% in the proportion of cancers that remained undiagnosed.

DM compared to DBT in one or two views: diagnostic accuracy

Six studies (mostly using a range of DBT-capable GE units or prototypes) explored the diagnostic performance of DM (not further described) compared to different combinations of DBT alone (i.e., either one view or two view and not in combination with FFDM or DM).

DBT (in one or two views) compared to DM resulted in increases in diagnostic accuracy which were similar to the increases reported above for FFDM + DBT and FFDM + DBT_{MLO}. In summary, the AUC measurement range with DBT_{MLO/CC} (with or without DM_{CC}) was 0.852 to 0.916. While the same direction of effect was reported for sensitivity (range 76% to 82% with DM compared to DBT in one or two views), the overall values were slightly lower compared to FFDM + DBT and FFDM + DBT_{MLO}. Reported specificity was very high in the two studies that reported specificity results (range: 85% to 95%). For DBT in one or two views compared to FFDM/DM, the AUC measurement range was 0.769 to 0.881; sensitivity was about 73%; specificity ranged from 86% to 90%. These results suggest DBT can be used in a range of different combinations to improve overall diagnostic accuracy.

Retrospective studies for women recalled to assessment

Haq et al. (2015) also used TOMMY trial data to consider whether DBT in two views is needed if the abnormality is seen in only one DM projection (see above for Gilbert et al.'s results comparing FFDM to FFDM + DBT and DBT + s2DM). Haq et al. compared DBT, FFDM, and ultrasound findings in their retrospective review of 617 possible masses, AD and asymmetries. While AUC was not reported, the authors found that on 1v-DBT (in the same projection as seen on FFDM) detected 84% of cancers and 42% were seen on both DBT views. The authors concluded that 1v-DBT may be adequate for the assessment of possible screen-detected soft-tissue abnormalities seen only on one FFDM view.

Retrospective studies for symptomatic women only

Bian et al. (2017) reported on DBT's diagnostic accuracy for the detection of masses in 631 Chinese women with dense breasts (48.8% had BIRADS c dense breasts, 51.2% had BIRADS d dense breasts) and who had undergone work-up for a symptomatic presentation. Using Selenia Dimensions, the authors compared FFDM to DBT. Bian et al. reported that 330 lesions were malignant and that DBT had a higher overall detection rate compared to FFDM: 84.3% compared to 77.3% ($p < .01$). Sensitivity for the detection of masses in women with dense breasts was higher with DBT (68.1% compared to 58.8%) as was specificity (95.2% compared to 86.7%).

In Kang et al.'s study, four reading arms (using three readers) were compared:

- FFDM
- DM_{CC} + s2DM_{MLO}
- FFDM + DBT_{MLO} (discussed above), and
- DM_{CC} + DBT_{MLO} + s2DM.

In this study, the addition of DBT did not result in a statistically significant increase in diagnostic accuracy. Mean JAFROC FOM for $DM_{CC} + DBT_{MLO}$ (0.907) was higher than FFDM alone (0.881) and $DM_{CC} + s2DM_{MLO}$ (0.848). The authors reported that FFDM + DBT_{MLO} had highest average JAFROC FOM across three readers ($p=.016$) and the highest sensitivity, and comparable specificity compared to the three other reading arms but there was not much difference between the three DBT arms. FFDM had the lowest scores on all these dimensions (averaged across all three readers). All three radiologists' diagnostic performance improved with DBT. Radiation dose was lower, which influenced Kang et al.'s conclusions that $DM_{CC} + DBT_{MLO}$ offered the best balance between improvements in diagnostic accuracy and the lowest mean glandular dose (MGD) increment (see discussion in *section 3.6*).

Retrospective studies: diagnostic pathway unknown

Chan et al. (2017) compared FFDM to DBT (wide angle 60°). From this cancer-enriched study of 134 women recalled to assessment or presenting symptomatically and using a fully crossed multi-reader design, AUC measurements for the probability of malignancy were presented. For all readers, the AUC measurement was increased with DBT (range = 0.884 to 0.936) compared to FFDM (range = 0.723 to 0.886), with statistically significant results reported for five out of six readers. An overall average increase was calculated (0.129; 95% CI: = 0.069, 0.188; $p < .05$). Overall sensitivity and specificity results were not presented but the authors noted that 17.8% more cancers would be considered suspicious with DBT compared to FFDM alone. A small increase in false positives (2%) with DBT was also reported. Again, as consistent with other studies, diagnostic performance improved with DBT.

Gennaro et al. (2013) recruited 250 women with at least one BIRADS > 3 lesion detected on FFDM or ultrasound and reported that mean AUC measurement over six readers for DM_{CC} + DBT_{MLO} were higher (0.852) than FFDM (0.831). The increased difference was not statistically significant ($p = .197$) but was within the non-inferiority range. The authors concluded that DM_{CC} + DBT_{MLO} improved diagnostic performance compared to FFDM and that this result was better than DBT_{MLO} alone. DBT was superior in terms of finding and characterising lesions and had an overall higher lesion detection fraction (66.5% with DM_{CC} + DBT_{MLO} compared to 60% with FFDM).

Endo et al. (2018) investigated a new method of creating a similar quality image to a conventional DBT image with lower radiation and still retaining the same number of views. This new technology involves three new DBT technologies (iterative reconstruction to suppress reflection of anything not in the focal plane, super-resolution to improve visibility of microstructures, granularity improvement to extract structure-less noise). It also includes two new FFDM technologies (pattern recognition technology and a sharpness improving technique). This new method of reconstruction and new processing (described as DBT (nd) and FFDM (nd)) was compared to conventional FFDM. This study is included in the literature review because it demonstrated the further development of DBT as a technology and highlights the fact that manufacturers have retained a strong focus on the adaptation and improvement of DBT as a technology. Using the newly developed technology, described as FFDM (nd) + DBT (nd), Endo et al. reported improvements in diagnostic accuracy with FFDM (nd) + DBT (nd) compared to conventional FFDM in a study of 913 Japanese patients. Statistically significant increases in overall AUC and sensitivity were reported (difference: 2.6, $p = .015$; difference: 5.1%, $p = .049$ respectively). There was a non-significant increase in specificity (difference: 2.8%, $p = .52$). Statistically significant increases in sensitivity were reported for all readers but mixed results were reported for specificity (although the difference in specificity was not significant). The authors reported that the newer technology, overall demonstrated improved diagnostic accuracy compared to conventional FFDM.

Overall, these results relate to different DBT-capable units and show promising evidence about future ways in which DBT could be optimised to assess women recalled to assessment for a mammographic abnormality and which could reduce the radiation exposure, although further studies in larger populations are needed to validate the results.

None of the studies reported PPV or NPV results.

s2DM compared to FFDM: diagnostic accuracy

s2DM images are generated from a DBT dataset and a separate FFDM image may not need to be acquired (that is, women only undergo the DBT examination without the DM imaging; bilateral

imaging with DBT + s2DM effectively requires only four projections compared to eight projections needed with FFDM + DBT). In their literature review, Michell & Batohi (2018) provided a short summary of the evidence underpinning the performance of s2DM compared to FFDM, citing many of the studies discussed in *Allen + Clarke's* literature review on DBT in screening and this literature review. Their conclusions were that emerging evidence suggests that s2DM (using Hologic's C-view software) is not inferior to FFDM imaging, that s2DM reduces overall radiation dose, and s2DM is a technique that is improving all the time. It is still uncertain whether s2DM's clinical performance is sufficient to replace FFDM as further large prospective trials are needed.

While not directly comparing DM and DBT, five studies (Garayoa et al., 2018; Mariscotti et al., 2017; Choi et al. 2016b; Kang et al., 2016; Gilbert et al., 2015) compared s2DM to FFDM. These studies are included because they provide advice on whether s2DM can be used as a single primary imaging tool that provides equivalent diagnostic performance compared to FFDM, but which reduces radiation dose (i.e., these studies support optimisation of DBT's performance in the assessment and diagnosis of breast cancer). s2DM's performance is reported as being at least equivalent to FFDM in all studies.

Retrospective studies for women recalled to assessment and symptomatic women

Garayoa et al. (2018) completed a large blinded retrospective study of 2384 images (bilateral CC and MLO views) to assess non-inferiority of s2DM to FFDM. This study, powered to establish non-inferiority of s2DM, found that s2DM was comparable to FFDM for malignancy probability and lesion visibility (mean AUC measurement: 0.811 and 0.812), and had an equal specificity and slightly lower sensitivity for s2DM. Using s2DM (reconstructed using Hologic's C-view software) instead of FFDM would result in a much lower radiation dose (up to 45% less) but still achieve comparable diagnostic performance, suggesting that s2DM reconstructed from DBT could replace FFDM if using the Hologic system.

Retrospective studies of women recalled to assessment

The remaining two studies looking at diagnostic performance of s2DM also reported equivalence/non-inferiority between s2DM and FFDM. Choi et al. (2016a) reported on a specific population group (women with small, invasive and often FFDM-occult T1 cancers), finding that for all three readers sensitivity and specificity increased with s2DM (used without DBT images) for women with both dense and less dense breasts. Increases were usually in the range of a couple of percentage points, but results were not significant. Earlier detection of these cancers could reduce interval cancer rates and increased specificity may result in fewer recalls to further imaging with s2DM compared to FFDM. In the final study (the TOMMY trial), Gilbert et al. (2015a) also investigated DBT + s2DM (study described above). The authors reported that DBT + s2DM significantly increased diagnostic accuracy (for AUC measurement, sensitivity and specificity) compared to FFDM (0.88 compared to 0.84) and was comparable to FFDM + DBT (0.88 compared to 0.89). More data from the TOMMY trial is reported above.

Retrospective studies of women presenting for screening and symptomatic women

Tagliafico et al. (2017) reported results of a sub-study from the ASTOUND trial which investigated the accuracy and reading time burden of using DBT as an adjunct to FFDM in screening. Six different strategies were investigated which varied in terms of modalities used and number of reads. The accuracy of reader performance was measured using ROC analysis which was used to estimate AUC. While all AUCs sensitivities, and specificities were all high having DBT + s2DM was found to be significantly better than all five other strategies (AUC measurement 0.979). Unsurprisingly, interpretation of DBT alone or s2DM alone (even when

read by two readers) was found to be the least accurate. These results came from a subset of a large prospective study, still the patient data included was relatively not that small (n=163).

Retrospective studies for symptomatic women only

Kang et al. (2016) reported on s2DM (either with DM_{CC} or DM_{CC} + DBT_{MLO} + s2DM). The authors reported that the reading arms that included s2DM (i.e., DM_{CC} + s2DM_{MLO} or DM_{CC} + DBT_{MLO} + s2DM_{MLO}) had similar diagnostic accuracy in terms of sensitivity and specificity results for their paired arm (i.e., FFDM or FFDM + DBT_{MLO}). Specificity was increased (but not significantly so).

Retrospective studies: diagnostic pathway unknown

Similar findings were reported in Mariscotti et al.'s recent retrospective analysis of a dataset of 231 women who had an imaging abnormality detected at screening (approximately half of which were suspicious microcalcifications). In this study, s2DM was compared to FFDM (using Hologic and C-view) (Mariscotti et al., 2017). Study results demonstrated that, overall, s2DM had a similar performance compared to FFDM in terms of AUC measurement as demonstrated by the shape of the ROC curve. BIRADS agreement for malignant and benign lesions presenting as masses or microcalcifications were also similar. Mariscotti et al. (2017) reported that sensitivity was increased with s2DM but that s2DM and FFDM had similar specificity. In this study, s2DM was not a superior test by itself, except for the visualisation of AD. The authors suggested that s2DM could be used (with DBT) for the assessment of all lesion types without the need to acquire a separate FFDM image given the comparable performance between FFDM and s2DM. This would result in a lower radiation exposure for women.

Table 6: AUC measurement, sensitivity, specificity and PPV/NPV results comparing DM to DBT

6A Prospective studies comparing FFDM to FFDM + DBT

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Mercier et al. (2015)	Prospective reader (n=2) analysis Participants: 74 French women and one man who had diagnostic mammography, DBT, ultrasound and MRI for staging of BIRADS 4 and 5 lesions (124 histology-proven malignant lesions) Unit: Siemens Mammomat Inspirations	No data	DM: 59.5 DBT: 75 Difference: $p=.00013$ Ultrasound: 79 MRI: 92.5	DM: 81.6 DBT: 74.4 Difference: $p=.27$ Ultrasound: Not stated MRI: Not stated	PPV DM: 53 DBT: 66.7 NPV DM: 81.6 DBT: 81.6
Alakhraas et al. (2014)	Prospective reader (n=26) performance study Participants: 50 cases (most of which were BIRADS 3) imaged with bilateral FFDM + DBT: 32 malignancies (20 IDC, 3 ILC, 5 DCIS, 4 others); 23 benign Unit: Hologic Selenia Dimensions	AUC ROC FFDM: 0.681 FFDM + DBT: 0.788 Wilcoxon test statistic: -337, $p<.0001$ JAFROC FFDM: 0.621 FFDM + DBT: 0.745 Wilcoxon test statistic: -274 $p<.00001$	FFDM: 63 FFDM + DBT: 70.4 Wilcoxon test statistic: -188, $p=.0011$	FFDM: 65.2 FFDM + DBT: 78.3 Wilcoxon test statistic: -237, $P=.0015$	No data
Cai et al. (2015)	Prospective case analysis Participants: 79 Chinese women (mean age = 41.76±8.75y) with pathologically-proven breast cancer (cohort=832) Unit: Hologic Selenia Dimensions	FFDM: 0.805 FFDM + DBT: 0.914	FFDM: 60 FFDM + DBT: 82.9	FFDM: 93.2 FFDM + DBT: 93.2	No data
Mansour et al. (2014)	Prospective reader (n=3) analysis blinded to pathology and other readers Participants: 166 consecutive Egyptian women (68.7% with dense breasts, BIRADS 3,4) recalled to assessment (n=72) or symptomatic (n=94), imaged with FFDM + DBT + ultrasound Unit: Hologic Selenia Dimensions	No data	FFDM: 60 FFDM + DBT: 94.5	FFDM: 20.7 FFDM + DBT: 74	PPV FFDM: 62 FFDM + DBT: 92 NPV FFDM: 20 FFDM + DBT: 80

6B Retrospective studies comparing FFDM to FFDM + DBT

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Mariscotti et al. (2016)	Retrospective, multi-reader (n=12) study Participants: 83 consecutive women with newly diagnosed ILC (107 ILC proven at histology) Unit: Hologic Selenia Dimensions	<i>Pooled from 12 readers</i> FFDM: 0.84 FFDM + DBT: 0.89, $p<.0001$	FFDM: 70 FFDM + DBT: 85, $p<.0001$	FFDM: 89 FFDM + DBT: 87, $p=.5$	No data
Seo et al. (2016)	Retrospective reader (n=6) evaluation (data collected prospectively) Participants: 203 symptomatic women or recalled from mammography or ultrasound (cohort=219); 98 had a previous breast cancer; 206 lesions (129 malignancies, 77 benign); exclusions (n=16) because of previous biopsy or surgical clip Unit: Hologic Selenia Dimensions	<i>JAFROC FOM</i> FFDM: 0.775 DBT alone: 0.807 FFDM + DBT: 0.827 <i>Difference (FFDM/FFDM + DBT): 0.052, $p<.001$</i>	FFDM: 73.2 DBT alone: 78.3 FFDM + DBT: 80.0 <i>Difference (FFDM/FFDM + DBT): 0.135, $p<.001$</i>	FFDM: 61.0 DBT alone: 63.0 FFDM + DBT: 64.3 <i>Difference (FFDM/FFDM + DBT): 0.524, $p<.182$</i>	No data
Bansal & Young (2015)	Retrospective reader (n=4) analysis with blinding to final histology and patient history Participants: 103 consecutive Welsh women with subtle mammographic or mammographically normal by suspicious ultrasound (16 malignancies; 90 normal); key excluded patients: BIRADS M5; microcalcifications Unit: Hologic Selenia Dimensions	FFDM: 0.721 (0.662–0.905) FFDM + DBT: 0.901 (0.765–1.00) <i>Difference: 0.180</i>	No data	No data	No data
Gilbert et al. (2015)	Retrospective, multicentre (n=6), blinded multi-reader (n=26) matched study (TOMMY trial); cancer-enriched cohort; prospective data collection Participants: 7060 women recalled to assessment or aged <50y with a family history of breast cancer (cohort=8869); exclusions not described; 1137 malignancies (788 IDC; 109 ILC; 203 DIC; 59 other); 5691 normal Unit: Hologic Selenia Dimensions and C-view software <i>TOMMY trial: also cited in Phi et al. (2018)</i>	FFDM: 0.84 (0.82, 0.86) FFDM + DBT: 0.89 (0.87, 0.91) s2DM + DBT: 0.88 (0.86, 0.90) Difference between FFDM + DBT and s2DM + FFDM compared to FFDM: $p<.001$	FFDM: 87 FFDM + DBT: 89 s2DM + DBT: 88	FFDM: 58 FFDM + DBT: 69 s2DM + DBT: 71	No data
Rafferty et al. (2014)	Multisite retrospective reader (n=15 with additional training in DBT but supplementary training) study of DM then DM + DBT _{MLO} then DM + DBT using a cancer enriched dataset Participants: 310 cases (51 biopsy proven malignancies)	DM: 0.828 DM + DBT _{MLO} : 0.864 DM + DBT: 0.895 <i>Differences in AUC</i>	<i>Overall</i> DM: 62.7 DM + DBT _{MLO} : 71.4 DM + DBT: 78.7	DM: 86.2 DM + DBT _{MLO} : 86.0 DM + DBT: 84.5	<i>PPV</i> DM: 47.3 DM + DBT _{MLO} : 50.2 DM + DBT: 50.1

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
	Unit: research prototype by Hologic <i>NB Study 2</i>	DM/DBT _{MLO} : 0.36 (0.009, 0.063), <i>p</i> =.009 DM/DBT: 0.068 (0.041, 0.095), <i>p</i> =.001 <i>Calcified lesions</i> FFDM: 0.817 FFDM + DBT: 0.831, <i>p</i> =.082 <i>Non-calcified lesions:</i> FFDM: 0.842 FFDM + DBT: 0.93, <i>p</i> <.001	<i>For in situ cancers</i> DM: 67.5 DM + DBT _{MLO} : 68.8 DM + DBT: 70.8 <i>For invasive cancers</i> DM: 60.6 DM + DBT _{MLO} : 72.6 DM + DBT: 82.3		<i>NPV</i> DM: 92.1 DM + DBT _{MLO} : 93.8 DM + DBT: 95.3
Rafferty et al. (2013)	Multisite retrospective reader (n=12 with no experience in DBT) of FFDM compared to FFDM + DBT Participants: 312 cases (48 biopsy-proven malignancies) (cohort) Unit: research prototype by Hologic <i>NB Study 1</i> <i>Also cited in Garcia-Léon et al. (2015)</i>	FFDM: 0.821 FFDM + DBT: 0.894 <i>Difference: 0.072 (0.037, 0.108), p</i> <.001 <i>Calcified lesions</i> FFDM: 80.4 FFDM + DBT: 84.0, <i>p</i> =.073 <i>Non-calcified lesions:</i> FFDM: 80.8 FFDM + DBT: 91.2, <i>p</i> <.001	FFDM: 65.5 FFDM + DBT: 76.2	FFDM: 84.1 FFDM + DBT: 89.2	No data
Waldherr et al. (2013)	Retrospective blinded reader (n=2) study Participants: 144 women either recalled to assessment or with symptoms; 86 malignancies Unit: Hologic Selenia Dimensions prototype <i>Also cited in Garcia-Léon et al. (2015)</i>	No data	FFDM: 74.4 (63.9, 83.2) DBT alone: 88.4 (79.7, 94.3) FFDM + DBT: 91.9 (84.0, 96.7)	FFDM: 76.8 (63.6, 87.0) DBT alone: 78.9 (66.1, 88.6) FFDM + DBT: 75.4 (62.2, 85.9)	<i>PPV</i> FFDM: 83.1 (72.9, 90.7) DBT alone: 86.4 (77.4, 92.8) FFDM+DBT: 84.9 (75.03, 91.5) <i>NPV</i> FFDM: 66.2 (53.4, 77.4) DBT alone: 81.8 (69.1, 90.9) FFDM + DBT: 86.0 (73.3, 94.2)

6C Retrospective studies comparing FFDM/DM to DBT alone (in one or two views)

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Chan et al. (2017)	Fully crossed, retrospective reader (n=6) study comparing DBT to FFDM in a cancer enriched dataset Participants: 134 cases recalled to assessment or with clinical findings, excluding microcalcification cases (61 malignancies, 85 benign/normal) Unit: GE GEN2 prototype (wide angle 60°)	<i>Mean AUC and BIRADS rating</i> FFDM: 0.769 DBT: 0.916 <i>Difference 0.129, p=.0001</i>	Not provided	Not provided	Not provided
Bian et al. (2016)	Retrospective reader (n=3) study Participants: 631 Chinese women with symptoms of breast cancer (330 malignancies) Unit: Hologic Selenia Dimensions	No data	FFDM: 68.1 DBT: 58.8 <i>False negative rate</i> FFDM: 26.6 DBT: 17.7	FFDM: 86.7 DBT: 95.2	No data
Kang et al. (2016)	Retrospective, single institution reader (n=3) study using prospectively collected patient images using four arms Participants: 130 women presenting for screening (72) or diagnosis (58); 159 lesions (27 histology-proven malignancies; 132 benign) Unit: GE SenoClaire with Volume preview software	<i>Mean JAFROC</i> FFDM: 0.881 DM _{CC} + s2DM _{MLO} : 0.848 FFDM + DBT _{MLO} : 0.914 DM _{CC} + DBT _{MLO} + s2DM: 0.907 Significance not achieved for any comparison	FFDM: 73 (55, 86) DM _{CC} + s2DM _{MLO} : 69 (52, 82) DM _{CC} + DBT _{MLO} : 83 (71, 90) DM _{CC} + DBT _{MLO} + s2DM: 82 (67, 90)	FFDM: 90 (86, 93) DM _{CC} + s2DM _{MLO} : 95 (92, 97) DM _{CC} + DBT _{MLO} : 94 (90, 96) DM _{CC} + DBT _{MLO} + s2DM: 95 (92, 97)	No data
Taha Ali et al. (2016)	Retrospective reader (n=3) study at a single institution Participants: 132 Egyptian women with at least one lesion (67 malignancies: 34 IDC, 17 ILC, 7 IDC + DCIS; 9 others; 78 benign) detected with DM +/- ultrasound Unit: GE SenoClaire	No data	FFDM: 55.2 DBT: 91	FFDM: 62.8 DBT: 92.3	<i>PPV</i> FFDM: 56 DBT: 91 <i>NPV</i> FFDM: 62 DBT: 92.3
Thomassin-Naggara et al. (2015)	Retrospective evaluation comparing FFDM to FFDM + DBT _{MLO} Participants: 155 symptomatic women (cohort=202) undergoing DM with subsequent percutaneous biopsy; 61 malignancies; 90 benign (histology-proven) Unit: Hologic Selenia Dimensions	FFDM: 0.692 FFDM + DBT _{MLO} : 0.809, <i>p<.01</i>	<i>Mean sensitivity</i> FFDM: 72.9 FFDM + DBT _{MLO} : 89.3	<i>Mean specificity</i> FFDM: 50.5 FFDM + DBT _{MLO} : 50.7	<i>PPV</i> FFDM: 48.9 FFDM+DBT _{MLO} : 53.4 <i>NPV</i> FFDM: 74.2

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
					FFDM+DBT _{MLO} : 88.2
Gennaro et al. (2013)	Retrospective reader (n=6) study Participants: 250 asymptomatic or symptomatic diagnostic patients (469 breasts); 68 malignancies, 401 benign/normal, patients with prior mastectomy and high family risk excluded Unit: GE prototype based on Senographe platform	<i>Mean AUC</i> FFDM: 0.831 DBT _{MLO} + DM _{CC} : 0.852	<i>Mean sensitivity</i> FFDM: 72.8 DBT _{MLO} + DM _{CC} : 76.2 <i>Difference: 3.4, p=.269</i>	<i>Mean specificity</i> FFDM: 83 DBT _{MLO} + DM _{CC} : 84.9 <i>Difference: 1.8, p=.130</i>	No data

6D Retrospective studies comparing FFDM to s2DM (reconstructed from FFDM + DBT in combo mode)

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Garayoa et al. (2018)	Retrospective observational single-site, blinded reader (n=3) study Participants: 2384 images (298 patients) attending for a screening or diagnostic mammogram (119 malignancies, 350 normal, 125 benign) and selected randomly or to ensure dataset had range of mammographic findings presenting at a normal breast imaging clinic; 54 patients included twice to support inter- reader variability sub-study Unit: Hologic Selenia Dimensions in combo mode	<i>Mean BIRADS AUC for malignancy probability</i> FFDM: 0.867 (0.835, 0.899) s2DM: 0.853 (0.826, 0.881) <i>Lesion visibility</i> s2DM: 0.811 (0.800, 0.823) FFDM: 0.812 (.800, .823)	<i>Mean for malignancy (BIRADS 4-5)</i> FFDM: 83 s2DM: 79	<i>Mean for benign and normal lesions (BIRADS 1-3)</i> s2DM: 88 FFDM: 88	No data
Mariscotti et al. (2017)	Retrospective reader analysis (n=2) of s2DM and FFDM images Participants: 231 Italian women (cohort = 4030) with at least one mammographic abnormality from which there were 250 biopsy suspicious breast lesions (consecutive dataset); patients with implants excluded Unit: Hologic Selenia Dimensions in combo mode	<i>Overall</i> FFDM: 0.80 s2DM: 0.83, p=.21 <i>Data by mammographic finding in Tables 11-14</i>	<i>Overall</i> FFDM: 87 s2DM: 92 <i>Data by mammographic finding in Tables 11-14</i>	<i>Overall</i> FFDM: 62% s2DM: 60% <i>Data by mammographic finding in Tables 11-14</i>	No data
Choi et al. (2016a)	Retrospective case control reader (n=3) study Participants: 107 consecutive patients (cohort = 826) with unilateral T1 cancer detected at screening, no clinical symptoms; 107 negative patients from same cohort (matched	No data	<i>Overall for three readers</i> FFDM: 60.7, 71.0, 70.1 s2DM: 62.6, 71.0, 71.0	<i>Overall for three readers</i> FFDM: 94.4, 72.9, 88.8 s2DM: 96.3, 84.1, 91.6	No data

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
	for age and breast density); DCIS excluded Unit: Hologic Selenia Dimensions in combo mode				

6E Retrospective studies comparing FFDM to s2DM (reconstructed from FFDM + DBT in combo mode)

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Endo et al. (2018)	Retrospective reader (n=8) study using a new FFDM + DBT technology Participants: FFDM (nd) 263 Japanese women recalled to assessment or symptomatic; FFDM (conventional): 650 from same cohort (913 in total); 129 malignancies; Unit: Amulet Innovality (FUJIFILM)	FFDM: 0.883 FFDM (nd): 0.888 FFDM (nd) + DBT (nd): 0.909	FFDM: 80.7 FFDM (nd): 80.7 FFDM (nd) + DBT (nd): 85.4	FFDM: 88.4 FFDM (nd): 86.9 FFDM (nd) + DBT (nd): 89.6	No data

3.2. Diagnostic performance of DBT compared to DSCV

DSCV (also called compression mammogram, spot view, cone view or focal compression) involves compression of a specific area of the breast. DSCV is useful for reducing anatomical noise, increasing visibility of breast architecture and is useful in assessing masses, asymmetries, architectural distortion and retroareolar presentations and suspicious areas close to the chest wall. To acquire accurate DSCV, compression paddles must be positioned precisely. If the lesion is not in the compression paddle field, it may not be seen. False-negative results are one of the issues affecting the use of DSCV in breast abnormality work-up. DSCV may be used with/without magnification.

Six studies reported on DBT's accuracy compared to DSCV (with or without magnification). All but one of these studies were based on women recalled to assessment. This section discusses studies in which DBT has been evaluated as a replacement for DSCV (i.e., the imaging protocol is DBT compared to DSCV rather than as adjunctive work-up such as FFDM + DSCV + DBT).

DSCV (with or without magnification) is used to evaluate and characterise suspicious abnormalities detected with DM in a screening setting or in the workup of a symptomatic patient. DBT can replace DSCV in the assessment of screen-detected soft tissue abnormalities due to improved margin delineation and better ability to confirm benign lesions. It also does not rely on the precise positioning required by DSCV making it easier to acquire accurate images.

The United Kingdom's breast screening program (NHSBSP) position statement on the use of DBT states that DBT:

"is at least as good as spot compression views for the assessment of possible soft tissue abnormalities... studies found no difference between spot compression views and tomosynthesis in the detection of calcifications...a trial of the GE Healthcare tomosynthesis system carried out in Nottingham and Derby, showed equivalence of tomosynthesis and spot compression views in the assessment of soft-tissue abnormalities...there is sufficient evidence to justify the use of the Hologic Dimensions tomosynthesis system and the GE Healthcare Essential (SenoClaire) tomosynthesis system in assessment"(Borelli and Oduko, 2016).

The BSA position statement on DBT relies on Borelli & Oduko's work for its position on the use of DBT in assessment. *Allen + Clarke's* literature review does not further discuss the findings from the literature underpinning either the NHSBSP statement or the BSA position statement (i.e., Cornford et al., 2016; Morel et al., 2014; Zuley et al., 2013; Michell et al., 2012) although study results are summarised in *Table 7*. Instead, it presents findings from more recent studies (particularly those published since 1 January 2016) and specific additional findings not clearly reported in the systematic reviews. Study results comparing DBT and DSCV are summarised in *Table 7*.

The six studies discussed in this literature review are listed overleaf. Summaries of study results are included in *Table 7 (A-D)*. Further information about the implementation considerations of this approach are discussed in *Chapter 6*.

Systematic reviews and meta-analysis

None

RCTs and prospective studies

RCTs: none

Two prospective studies:

Women recalled to assessment only: Tagliafico et al. (2012)

Symptomatic women only: Noroozian et al. (2012)

Retrospective studies

Four retrospective studies:

Women recalled to assessment only: Heywang-Köbrunner et al. (2017); Whelehan et al. (2017); Cornford et al. (2016), Lockie et al. (2014, oral abstract only); Brandt et al. (2013)

Position statements

One statement: Borelli and Oduko (2016)

Key findings

Data from six studies comparing DBT to DSCV reported very high AUC measurement and sensitivity for both imaging protocols. While DBT's equivalence or superiority was not demonstrated through statistical testing, the differences in AUC and sensitivity were small and usually favoured DBT. That is, AUC measurements for DBT ranged from 0.87 to 1 compared to 0.857 to 0.963 for DSCV. Reported sensitivity was also high: DBT's sensitivity was higher than or equivalent to DSCV, with DBT values ranging between 96.9% and 100% compared to 85.6% to 100%. Mixed results and wider values ranges were reported for specificity: studies reported either higher specificity with DSCV or equivalence. Data from the Maroonah study indicated that improved accuracy with DBT (compared to DSCV) would result in fewer biopsies and other imaging work-ups such as ultrasound. Reported diagnostic equivalence or superiority for DBT compared to DSCV has already resulted in practice changes including in screening programs like the BSA and the NHSBSP (in locations where DBT-capable units are available) as well as in individual clinical practices.

Systematic reviews and narrative literature reviews

While some of the systematic reviews included papers comparing DBT to DSCV, none of these reported specifically on the diagnostic performance in the evaluation of suspicious lesions.

Prospective studies of women recalled to assessment

AUC measurement, sensitivity and specificity values from two prospective studies (both discussed in Garcia-León et al.'s systematic review, methodology discussed previously). Noroozian et al. (2012) and Tagliafico et al. (2012), in two small prospective studies using cancer-enriched cohorts completed when DBT was an emerging technology and using different DBT-capable units, both reported non-significant increases in AUC measurement (i.e., results favoured DBT).

Tagliafico et al. consecutively recruited 52 recalled women (nine cancers) presenting with non-calcified suspicious BIRADS 0 lesions. The authors used DBT and DSCV to assess two sets of

images from a single patient and reported a non-statistically significant increase in diagnostic accuracy when DBT was used (DBT = 1 compared to 0.963 for DSCV). The authors reported sensitivity of 100% for both DBT and DSCV although reporting such a high rate may have been affected by the exclusion of women with a microcalcification-based presentation from the study. Reporting higher results for PPV/NPV (that is, better ability to discriminate between malignancy and normal/benign findings) with DBT compared to DSCV led these authors to conclude that the diagnostic accuracy of DBT was at least equivalent to DSCV (but significance was not achieved).

Prospective studies of symptomatic women

Noroozian et al.'s study involved 67 women with lesions classified as BIRADS 4 or 5 (30 cancers). Like Tagliafico et al. (2012), Noroozian et al. reported a small, non-significant increase in AUC but reported that a further seven cancers (and five benign lesions) would have been recommended for biopsy with DBT. This suggests that both increased sensitivity and decreased specificity (although no sensitivity, specificity or PPV/NPV data was reported in their study).

Retrospective studies of women recalled to assessment

Diagnostic performance results reported in three studies with a retrospective design published since 2013 and using commercially available DBT-capable units (although one early study used a Hologic prototype unit) indicated that DBT had similar but was not superior to DSCV.

Hewang-Köbrunner et al. (2017) and Whelehan et al. (2017) are two multi-reader studies performed using data from the same cohort of women to determine if DBT was inferior to DSCV. Imaging data was collected prospectively in Germany from women undergoing DSCV for suspected breast cancer, with DBT imaging supplemental to the DSCV imaging. In these studies:

- Heywang-Köbrunner et al. (2017) included 241 women and reported on a clinical assessment pathway for screen-detected abnormalities in which FFDM + ultrasound + wide angle 1-v DBT was compared to FFDM + ultrasound + additional mammographic views (not further specified but we assume this to be DSCV given the study description).
- Whelehan et al. (2017) included 230 women and reported on screening mammogram + 1v-DBT (not specified) compared to screening mammogram + DSCV.

Regarding findings for AUC measurement, the DBT protocol in Heywang-Köbrunner et al.'s study reported a slightly lower value for DBT compared to additional DM views: AUC was 0.889 with the DBT protocol compared to 0.903 for the DM protocol (not significant). The authors reported heterogeneity in reader performance for DBT which could have resulted in the trend towards lower AUC measurement for DBT: two of the readers with double reading caseloads reported lower AUC measurement than the other readers. Whelehan et al. reported that FFDM + DBT (one view) can be an equivalent test to FFDM + DSCV for non-calcified lesions. AUC measurement indicated equivalence with slight superiority favouring DBT (0.87 compared to 0.857, $p=0.4890$).

Regarding sensitivity and specificity, Heywang-Köbrunner et al. (2017) reported that overall sensitivity appears to improve when DBT is used in the assessment process (96.9%); however, they also reported high sensitivity with the DSCV as well (95.4%). Lower overall sensitivity was reported by Whelehan et al. (2017) but a larger (not significant) increase with DBT protocol was reported (see *Table 7*). Likewise, differences in specificity were also reported, with lower rates in Heywang-Köbrunner et al.'s study and higher rates in Whelehan et al.'s study. The differences between these two studies (even though they draw on almost the same cohort) are most likely due to differences between the two imaging protocols. Also, it is likely that the studies themselves may be underpowered to accurately detect sensitivity and specificity.

Preliminary results from the Maroondah trial were only identified in an oral abstract (Lockie et al., 2014). In this study, 557 women had DSCV (compression and cone views) and DBT. Lockie et al. (2014) reported that 83.2% of study participants had benign final findings and DBT would result in a decrease in biopsy of 30% compared to imaging work-up with DSCV (from 16.7% to 11.8%, $p=.03$). A 13% reduction in ultrasound use was also likely when DBT was used (from 58.0% to 50.1%, $p=.03$). There was a non-significant increase in PPV for DBT compared to FFDM (83% compared to 82%) and a non-significant increase in NPV (from 84% to 92%, $p=.06$). Lockie et al. concluded that DBT could be used to reduce biopsy.

While they did not report on AUC measurement, Brandt et al. (2013) reported that in women recalled to assessment for non-calcified abnormalities, diagnostic work-up required an average of three views. DBT provided adequate information with which to confirm next assessment steps without supplementary mammographic imaging in 93-99% of cases. The authors concluded that DBT could replace DM views for AD, asymmetries and masses. Sensitivity results were 100% for sensitivity for two out of three readers for DM and DBT although the authors noticed that one reader relied more heavily on ultrasound results to confirm final BIRADS classification. Given this and the lack of statistical difference in key study metrics, Brandt et al. concluded that DBT could replace conventional diagnostic mammography, especially because of the reduction in additional individual images required to complete diagnostic assessment (i.e., the DBT images compared to up to six mammographic and/or ultrasound images). Despite the clinical simulation approach, findings from Brandt et al. may not reflect actual clinical findings as a Hologic beta unit was used to acquire DBT images.

No data on PPV/NPV was presented in these studies.

Table 7: AUC measurement, sensitivity, specificity and PPV/NPV results comparing DBT to DSCV

7A Prospective studies comparing DBT to DSCV in symptomatic women

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Noroozian et al. (2012)	Prospective blinded reader (n=4) performance study comparing DBT (one or two views) to additional mammographic spot views (MSV) Participants: 67 women with BIRADS 4 or 5 lesions (30 malignant, 37 benign) Unit: combined GE ultrasound and DBT <i>Also cited in Garcia-Léon et al. (2015)</i>	MSV: 0.90 DBT: 0.91, $p=.60$	No data	No data	No data

7B Prospective studies comparing DBT to DSCV in women recalled to assessment

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Tagliafico et al. (2012)	Prospective study Participants: 52 consecutive women recalled to assessment with a BIRADS 0 mass, AD or focal asymmetries and imaged with both DBT and DSCV. 9 malignancies. Women with microcalcification were excluded. Unit: Hologic Selenia Dimensions <i>Also cited in Garcia-Léon et al. (2015)</i>	DBT: 1 DSCV: 0.963, $p=.43$	DBT: 100 (91, 100) DSCV: 100 (91, 100)	For both radiologists: DBT: 100 (91, 100) DSCV: 94 (91, 100)	PPV DBT: 100 DSCV: 82 NPV DBT: 100 DSCV: 100

7C Retrospective studies of women recalled to assessment

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Heywang-Köbrunner et al. (2017)	Multi-reader (n=16) study (FFDM + wide angle 1v-DBT + ultrasound compared to FFDM + additional mammography views + ultrasound); data collected prospectively in <i>Whelehan et al. (2017)</i> Participants: 241 cases (35 malignancies; 206 benign/normal); malignancies: 19 masses, 11 AD, 2 global distortion, 3 focal asymmetries; microcalcifications excluded Unit: Siemens Mammomat Inspiration <i>Data from Whelehan et al. (2017)</i>	DM protocol: 0.903 (0.886, 0.921) DBT protocol: 0.889 (0.871, 0.907).	<i>Average sensitivity</i> DM protocol: 95.4 DBT protocol: 96.9, $p=.38$ <i>BIRADS 4 and 5 only:</i> DM protocol: 95.1 DBT protocol: 96.0, $p=.66$	<i>Average specificity</i> DM protocol: 58.2% DBT protocol: 50.0%, $p<.001$ <i>BIRADS 4 and 5 only:</i> DM protocol: 63.0% DBT protocol: 52.5%, $p=.001$	No data provided
Whelehan et al. (2017)	Retrospective, multi-reader (n=8) study comparing FFDM + DBT to FFDM plus supplementary mammographic views Participants: 230 German women participating in a population-based screening program recalled to assessment for non-calcific abnormalities (20 malignant masses, 11 AD, 4 asymmetric densities, 195 benign cases) Unit: Siemens Mammomat Inspiration	FFDM + SMV: 0.857 FFDM + DBT: 0.870, Difference: 0.013, $p=.4890$	FFDM + SMV: 86 FFDM + DBT: 90	FFDM + SMV: 64 FFDM + DBT: 59, $p=.0002$	No data

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Brandt et al. (2013)	Retrospective, blinded reader (n=3) evaluation of DBT compared to conventional mammography views (range 1-6 views) in the workup of noncalcified asymmetries, AD, masses Participants: 146 women recalled to assessment with 158 abnormalities Unit: Hologic beta unit <i>Also cited in Garcia-Léon et al. (2015)</i>	No data	FFDM: 100 (68, 100) DBT Reader 1: 100 (68, 100) Reader 2: 100 (68, 100) Reader 3: 88 (53, 98)	FFDM: 94 (88, 97) DBT Reader 1: 94 (88, 97) Reader 2: 93 (87, 96) Reader 3: 89 (83, 94)	No data

7D Retrospective studies informing the NHSBSP position statement on tomosynthesis

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Cornford et al. (2016)	Dual paired comparison study with retrospective multi-reader analysis of diagnostic performance of DBT (using GE prototype DBT unit) compared standard supplementary views (spot compression) and prospective clinical utility of DBT in screening assessment Participants: 322 women recalled to assessment for soft tissue abnormalities.	Prospective results DM + DSCV: 0.922 DBT: 0.946 Retrospective results DM + DSCV: 0.873 DBT: 0.900	Prospective results <i>Absolute sensitivity (M5 lesions)</i> FFDM + DSCV: 44.24 FFDM + DBT: 57.52 $\chi^2=0.045$ <i>Complete sensitivity (M3,4,5)</i> FFDM + DSCV: 90.26 FFDM + DBT: 93.80 $\chi^2=0.326$	Prospective results FFDM + DSCV: 78.6 FFDM + DBT: 82.53 $\chi^2=0.288$	<i>PPV (M5)</i> FFDM + DSCV: 98.03 FFDM + DBT: 97.01 $\chi^2=0.726$ <i>NPV (M1/2)</i> FFDM + DSCV: 94.24 FFDM + DBT: 96.42% $\chi^2=0.307$
Morel et al. (2014)	Retrospective matched comparison reader (n=7) review study comparing diagnostic accuracy of FFDM + 1v-DBT (not specified) to FFDM + coned compression mammography (CCMM) Participants: 341 women (103/354 lesions were malignant and of these 80 were soft-tissue and 23 were DCIS) recalled for assessment following screening or symptomatic mammography Unit: Hologic Selenia Dimensions	Whole dataset: FFDM + CCMM: 0.87 (0.83, 0.91) FFDM + DBT: 0.93 (0.91, 0.95) Difference is 0.06, $p=.0014$ Soft-tissue lesions: FFDM + CCMM: 0.90 FFDM + DBT: 0.97, $p=.005$	Absolute sensitivity (M5): FFDM + CCMM: 41.74 FFDM + DBT: 52.42	Absolute sensitivity: Not reported Complete specificity (M3+M4+M5): FFDM + CCMM: 53.78 FFDM + DBT: 64.54	<i>PPV (M5)</i> FFDM + CCMM: 91.48 FFDM + DBT: 87.09 <i>NPV (M1)</i> FFDM + CCMM: 96.47 FFDM + DBT: 100

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Zuley et al. (2013)	Retrospective review by readers (n=8) comparing DBT to diagnostic mammography (two views, not further described) Participants: 217 consecutive lesions (72 malignancies in 182 women; 182 masses, 25 asymmetries, 10 AD) Unit: Hologic prototype	Average area under the probability of malignancy-based ROC curve DM images: 0.83 (0.77, 0.83) DBT: 0.87 (0.82, 0.92) Difference 0.04, $p<.001$	True positive fraction for BIRADS 3,4 or 5 DM images: 0.96 DBT: 0.96 True positive fraction BIRADS 4 or 5 DM images: 0.89 DBT: 0.90 True positive fraction BIRADS 5 DM images: 0.33 DBT: 0.39	False positive fraction for BIRADS 3,4 or 5: DM images: 0.85 DBT: 0.74, $p<.01$ False positive fraction BIRADS 4 or 5 DM images: 0.57 DBT: 0.48, $p<.01$ True positive fraction BIRADS 5 DM images: 0.02 DBT: 0.02	No data
Michell et al. (2012)	Retrospective study Participants: 738 women recalled to assessment (204 malignancies; 286 benign, 269 normal)	FFDM + SFM: 0.8949 FFDM + DBT + SFM: 0.9671 $p=.0001$	No data	No data	No data

3.3. Inter-observer agreement increases with the use of DBT

Inter-observer agreement is an important measure of the overall accuracy of data collected and assessed through a diagnostic test. It is usually reported as a kappa statistic.¹¹ Increased agreement in reporting of mammogram outcome between readers suggests that more readers are making the same assessment about the likelihood of malignancy in each image, leading to greater overall confidence in the reproducibility of reader' decisions.

One systematic review and twelve studies, while reporting on overall diagnostic accuracy, also reported on the level of inter-observer agreement when using DBT compared to different DM modalities. Most of these studies were smaller and retrospective. Studies are listed below.

Systematic reviews and narrative literature review

Systematic review: Garcia-Léon et al. (2015)

Narrative literature reviews: none

RCTs and prospective studies

RCTs: none

Prospective studies: none

Retrospective studies

12 retrospective studies:

¹¹ A kappa statistic value of 0.8–1.0 shows 'almost perfect' agreement, 0.6–0.8 represents 'substantial' agreement, 0.4–0.6 show 'moderate' agreement, and 0.2–0.4 show 'fair' agreement.

Only women recalled to assessment: Dibble et al. (2018); Choi et al. (2016a); Brandt et al. (2013); Kopans et al. (2011)

Women recalled to assessment and symptomatic women: Garayoa et al (2018)

Women presenting for screening and symptomatic women: Galati et al. (2017)

Only symptomatic women and/or women with breast cancer: Tang et al. (2017)

Diagnostic pathway unclear: Fornvik, et al. (2018); Mariscotti et al. (2017); Chan et al. (2017); Choi et al. (2016b); Clauser et al. (2016)

Key findings

Inter-observer agreement is an important measure of the overall accuracy of data collected to detect and evaluate breast lesions. One systematic review and 12 studies reported on inter-observer agreement or variation when using DBT compared to DM.

When DBT is used, inter-observer agreement about whether a lesion is benign or malignant increases

Based on smaller retrospective observer studies in a range of diagnostic groups, current evidence suggests an increase in inter-observer agreement for BIRADS classification with the use of DBT compared to DM. This increase in agreement was observed in the following DBT reading protocols: FFDM compared to FFDM + DBT, FFDM compared to DBT alone, or FFDM compared to s2DM. Inter-observer agreement (as measured by kappa statistic) with DBT imaging increased in all studies that reported it, with the use of DBT increasing overall agreement from moderate to good or excellent. Reported increases were usually about 25%. FFDM + DBT appears to be a more reliable test for BIRADS agreement with kappa statistics exceeding 0.8 with much lower rates reported for FFDM (ranging from 0.58 to 0.873). The use of s2DM compared to FFDM also appears to improve inter-observer agreement, suggesting that lesion conspicuity is greater with the reduction in 'noise' available with the 3D reconstruction algorithm.

There are significant increases in inter-observer agreement for subtle mammographic presentations like AD

Inter-observer agreement improved for subtle mammographic presentations like AD and asymmetries as did subjectively assessed reader confidence in detection of breast cancers with a subtle mammographic presentation. This is an important finding given the high PPV for malignancy that DBT-detected AD has (over 40%) and that AD is a primary mammographic finding characteristic of missed cancers. Reduction in inter-observer variability and increases in inter-observer agreement are likely to translate to higher quality clinical care.

3.3.1. There is some evidence from smaller, retrospective studies that when DBT is used, inter-observer agreement about whether a lesion is benign or malignant increases

Systematic reviews

Only one systematic review (Garcia-Léon et al., 2015, review methodology described previously) reported on inter-observer agreement. No pooled analysis was provided. Citing findings from Tagliafico et al. (2012), the authors only reported that DBT and DSCV resulted in a high degree of inter-observer agreement (kappa statistic = 0.95 for both DBT and DSCV, 95% CI: 86, 100). No further information is included in either Garcia-Léon or Tagliafico et al.'s papers.

Retrospective studies

Twelve studies reported on changes in inter-observer agreement (either for BIRADS agreement or recall agreement) using a range of FFDM/DBT imaging protocols. Studies were set in a wide variety of settings (as described above). There is moderate evidence that DBT (either alone or with FFDM, or s2DM) appears to be a more reliable test in terms of reducing inter-observer agreement and recall to assessment compared to FFDM. This literature review will only comment on study findings relating to inter-observer agreement (not decisions to recall asymptomatic women to assessment). Regardless of study diagnostic pathway, studies reported increases in inter-observer agreement (and therefore decreases inter-observer variation) of about 25%.

FFDM + DBT compared to FFDM: Inter-observer performance improved with DBT

Inter-observer agreement for BIRADS classification (i.e., the reported mammogram result) improved with FFDM + DBT compared to FFDM.

Reviewing data from women presenting for screening and the work-up of symptomatic women, Galati et al. (2017) compared FFDM to FFDM + DBT. In this study, two experienced readers (15 and 10 years' experience) reported increased agreement in BIRADS classification reported Cohen's kappa statistic of over 0.8 for FFDM + DBT with a much lower kappa statistic reported for FFDM (0.58). This indicates that FFDM + DBT is more reliable than FFDM alone.

DBT compared to FFDM: Inter-observer performance improved with DBT

Inter-observer variation in BIRADS classification was significantly reduced in Chan et al.'s 2017 study. They completed a multi-observer study in which six radiologists compared side-by-side DBT and FFDM projections before recording BIRADS evaluation of 142 malignant and benign soft-tissue lesions. Reported intraclass correlation coefficient between any two readers was between 0.435 and 0.679 for FFDM and rising to 0.767 to 0.837 for DBT. The intraclass correlation coefficient for BIRADS assessment also increased with the use of DBT (0.873, 95% CI 0.835, 0.904 with FFDM compared to 0.948, 95% CI 0.928, 0.963, $p < .0005$). The authors concluded that DBT results in greater inter-reader agreement compared to FFDM views for BIRADS classification of known cancers.

Brandt et al. (2013) approached the issue of inter-observer agreement in a slightly different way. In their retrospective observer study of 146 women recalled to assessment (158 non-calcified lesions: 99 asymmetries; 28 possible AD, 31 possible masses), three radiologists reviewed screening mammograms, diagnostic DM (not further described) and DBT images. Readers had access to previous mammogram and ultrasound images to assess the diagnostic accuracy of DBT compared to diagnostic mammography. Brandt et al. reported that there was good to excellent agreement between radiologists using DBT (with/without ultrasound) for non-calcified lesions compared to conventional DM views (with/without ultrasound), with over 90% agreement in BIRADS categories for all three readers (kappa statistic using Cicchetti-Allison method = 0.67 to 0.87). Specific information about BIRADS distribution was not provided for DBT images. It is important to note that this is a very early study with DBT imaging performed on a Hologic second generation beta unit, but it demonstrates that, even on older units, inter-observer agreement improved.

Subtle mammographic presentations (like AD) can be challenging to detect, resulting in larger inter-observer variation and decreased program efficacy. In Dibble et al.'s 2018 case-control study of women recalled to assessment, overall inter-observer agreement for a subtle mammographic presentation (AD) increased with the use of DBT (kappa statistic = 0.61) compared to FFDM (kappa statistic = 0.37). Two radiologists were experienced (with 9- or 19-

years' breast radiology) and two were completing their breast imaging fellowship. While both attending and fellow radiologists' agreement improved, greater improvement in agreement was seen for the more experienced radiologists (increase in kappa of 0.32 with DBT) compared to less experienced radiologists (increase in kappa of 0.23 with DBT). Increase in agreement was accompanied by an increase in reader confidence in the detection of possible AD for all radiologists (on a scale of 0-4, confidence increased from 2.6 with FFDM to 3.2 with DBT, $p=.001$). These are important findings given that AD is a subtle presentation: increasing consensus in detection is likely to improve quality of care through correct and timely diagnosis of cancers with this mammographic finding.

Choi et al. (2016b) completed a small ($n=35$ women with biopsy-proven breast cancer) retrospective observer performance study evaluating FFDM compared to different DBT imaging protocols, with DBT images obtained from a KERI prototype. Information about primary mammographic finding is not provided and mostly this study aims to demonstrate the equivalence of the KERI prototype to other DBT-capable systems. Radiologists were experienced (8 to 20 years breast imaging). Choi et al. reported fair to moderate or substantial agreement between readers for both FFDM and DBT views: lesion detectability (kappa statistic = 0.59 to 0.62); lesion location (kappa statistic = 0.52 to 0.84); lesion type (kappa statistic = 0.46 to 0.70) and BIRADS final assessment (kappa statistic = 0.48 to 0.69). In terms of detection, agreement when using DBT was superior to FFDM ($p<.046$). Given that this study was performed on a prototype unit and in the absence of further information about inter-reader agreement of lesion type and location and BIRADS final assessment, it is difficult to determine whether there was improvement with DBT compared to FFDM (which is shown in other studies).

Inter-observer performance from three other studies demonstrated a consistent effect with the studies described above (that is, higher agreement or less variation is seen when DBT is used). Studies reported only very brief results, which are described in *Table 8* (below). Two studies (Clauser et al., 2016; Kopans et al., 2011) reported on reader variation in the detection of microcalcifications. Kopans et al. reported similarities in performance whereas Clauser et al. reported significant differences. The other (Fornvik et al., 2018) is percent only.

Table 8: Additional studies reporting on inter-reader agreement/variation

Study	Study details	Results
Fornvik et al. (2018)	Retrospective reader ($n=2$) review Participants: 152 Japanese women with 103 malignancies (64 NST, 26 DCIS, 6 ILC, 4 mucinous carcinomas, 3 others) with 3T-MRI and FFDM + DBT in women with more dense breasts ($c=67.8\%$; $d=14.5\%$) Unit: Hologic Selenia Dimensions	<i>DBT</i> Reader 1: sensitivity 80.6%; PPV 76.1%; parametric AUC 0.875 Reader 2: sensitivity 82.5%; PPV 74.6%; parametric AUC 0.906
Clauser et al. (2016)	Retrospective blinded reader ($n=4$) study Participants: 150 FFDM and DBT cases (50 benign, 50 malignant, 50 BIRADS 1): women recalled to assessment or having diagnostic workup Unit: Siemens Mammomat Inspirations + Hologic Selenia Dimensions	Statistically significant differences in performance for detection of microcalcifications ($p<.003$)
Kopans et al. (2011)	Retrospective reader ($n=2$) unblinded paired comparison study to compare microcalcification conspicuity on FFDM compared to DBT _{MLO} Participants: 119 sequential cases: not clearly benign microcalcifications Unit: GE unit (unspecified, presumably prototype)	<i>Similar performance in the two modalities</i> Reader 1: 50/119 microcalcifications seen better on DBT; 9/119 seen better on FFDM Reader 2: 49/119 microcalcifications seen better on DBT; 10/119 seen better

Study	Study details	Results
		on FFDM; weighed kappa statistic: 0.9559

s2DM compared to FFDM: Inter-reader performance improved

Three studies looked at inter-reader variation comparing FFDM to s2DM.

In Garayoa et al.'s retrospective reader study, the three participating readers were experienced (i.e., assessing more than 5000 mammograms and 7000 DBT studies per annum; all had one year's experience reading s2DM images). Readers were blinded to clinical history and rated visibility of mammographic findings and provided a BIRADS classification for the most suspicious finding. All readers reported consistent BIRADS classification agreements for both imaging modalities (kappa statistic 0.70 for FFDM and 0.76 for s2DM). For main mammographic findings, kappa statistic was higher for s2DM for nodules (i.e., masses) and AD. Inter-reader agreement was lower for density and microcalcifications (but the study only had a small number of these which could explain the difference).

Similar increases in agreement were reported by Mariscotti et al. (2017): Cohens' kappa statistic = 0.93 for s2DM, no data presented for FFDM. Choi et al. (2016b) completed a retrospective study of the interpretive performance of s2DM and FFDM for T1 breast cancer, with increased agreement reported with s2DM for both final BIRADS assessment and visibility score (s2DM kappa statistic 0.651 and 0.472 compared to 0.548 and 0.366 for FFDM). The authors suggested that this was due to increased lesion conspicuity seen with s2DM. Even though it is a 2D image, the reconstruction process removes further noise creating a clearer image.

FFDM + DBT compared to FFDM +MRI: Inter-observer performance

Tang et al. (2017) (methodology described in *Part 8*) also reported excellent inter-observer agreement for BIRADS classification agreement between the two experienced radiologists in their study, reporting a Cohen's kappa statistic of 0.919 for FFDM + DBT compared to 0.894 for FFDM + MRI. No further analysis inter-reader agreement was available in this study.

3.4. Avoidance of unnecessary work-up for benign final outcome

The BIRADS lesion classification scheme is a way for radiologists to classify mammography findings for screening or assessment. BIRADS is widely used throughout North America and much of Europe. Use of BIRADS to classify mammogram results is different from the system used to classify breast density (which is also referred to as BIRADS density classification). BIRADS is not routinely used to report mammographic findings in the BSA program (which uses the NCBCC Synoptic Breast Imaging Report); however, it is the reporting system described in almost all the literature identified and reviewed as part of this report (the NCBCC reporting system was not reported on).¹²

¹²

BIRADS

0: incomplete/more imaging needed

1: negative

2: benign findings

3: probably benign findings

4: suspicious abnormality

5: highly suspicious abnormality

6: known biopsy with proven malignancy

Mammographic reporting and lesion classification are an important component of the assessment of a test's accuracy, but the findings presented in this section may be of limited applicability to the Australian context given that this system is not used in the BSA program, nor is short-term follow-up offered for inconclusive (BIRADS 3) results.

Accurate BIRADS lesion classification at assessment (enabled by clear lesion conspicuity, shape and margin) is an important way to ensure that further imaging required is minimised and that dismissal of a benign lesion or diagnosis of a cancer is confirmed quickly. Diagnostic mammogram and ultrasound results are evaluated and reported using systems like BIRADS and the Royal College of Radiologists' system. A mammogram is considered positive and requires further work-up if it is given a classification of BIRADS 3/M3 or above.

The implications of changing lesion classifications through increased accuracy of mammogram classification would be:

- BIRADS 0: a reduction in BIRADS 0 (incomplete examinations) means more complete/detailed image acquisitions
- BIRADS 1 or 2/M1 or 2: increases in BIRADS 1 or 2/M 1 or 2 demonstrate more cases being assessed as normal or benign lesions and requiring no further imaging or follow-up
- BIRADS 3/M3: decreases in BIRADS 3/M3 imply greater reader confidence and improved program accuracy to evaluate lesions as normal/benign or malignant, resulting in fewer cases of short interval follow-up imaging and fewer ongoing tests for women and greater peace of mind, and
- BIRADS 4 or 5/M4 or 5: increases in these classifications require more biopsy to confirm malignancy so a corresponding increase in PPV₃ is also important.

This literature review describes findings from one narrative literature review, four prospective studies and 16 retrospective studies that investigated changes between final BIRADS lesion classification when DBT is used compared to other mammographic views. Samples sizes varied between 31 women and more than 4500 participants, although most studies had fewer than 400 participants. These views were either repeated FFDM, unspecified views (i.e., DM is the only descriptor provided in the study) or DSCV. Information on changes to final BIRADS lesion classification by primary mammographic feature (i.e., AD, asymmetry, microcalcification, mass) is discussed in *section 3.5* of this report.

A total of 21 papers presented results on initial and/or final BIRADS lesion classification agreement comparing DBT to FFDM/DM. The studies discussed in this literature review are listed below/overleaf. Detailed study findings are presented in *Table 9 (A-C)*. Further findings

Royal College of Radiologists

M1: normal tissue

M2: benign

M4: probably malignant

M5: malignant

NCBCC Synoptic breast Imaging Report

1: no significant abnormality

2: benign findings

3: indeterminate/equivalent findings

4: suspicious findings of malignancy

5: malignant findings

from the ETOLE study may also provide additional evidence about the impact that viewing DBT images can have on final BIRADS lesion classification.

Systematic reviews and meta-analysis

Systematic reviews: none

One narrative literature review: Tingberg & Zackrisson (2011)

RCTs and prospective studies

RCTs: none

Four prospective studies:

Women recalled to assessment and symptomatic women: Bahrs et al. (2018); Mansour et al. (2014)

Only symptomatic women and/or women with a diagnosed breast cancer: Chae et al. (2016); Takamoto et al. (2013)

Retrospective studies

17 retrospective studies:

Only women recalled to assessment: Galati et al. (2017); McDonald et al. (2017); Lee et al. (2016); Haq et al. (2015); Lourenco et al. (2014); Morel et al. (2014); Brandt et al. (2013)

Only symptomatic women and/or women with a diagnosed breast cancer: Raghu et al. (2016); Taha Ali et al. (2016); Bansal & Young (2015); Mercier et al. (2015); Yang et al. (2013); Zuley et al. (2013)

Diagnostic pathway unclear: Chan et al. (2017); Kamal et al. (2016); Lång et al. (2014); Waldherr et al. (2013)

Key findings

The BIRADS lesion classification scheme is a way for radiologists to classify mammography findings. Accurate BIRADS lesion classification (enabled by clear lesion conspicuity, shape and margin) can ensure that unnecessary imaging is minimised, and that dismissal or diagnosis of suspicious lesions is confirmed quickly. Changes to reporting could be indicative of avoided procedures if a drop in inconclusive results – BIRADS 3 – are reported. A total of 19 studies compared BIRADS lesion classifications provided at DM imaging with DBT imaging.

Use of DBT results in a decrease in the number of inconclusive results compared to DM views (i.e., DBT contributes to the avoidance of additional follow-up or further work-up for lesions with a benign final outcome)

In the BSA program, mammogram results are reported using the NCBCC Synoptic Breast Imaging Report. No research identified in this report used the Australian reporting system. Almost all studies used BIRADS classifications for reporting mammogram results. Consistent findings were reported regardless of imaging protocol, study design, participants/study sub-population, or DBT unit used. Using DBT resulted in consistent changes to BIRADS lesion reporting. While not all results achieved statistical significance, all studies reported that:

- BIRADS 3 lesion reporting (i.e., inconclusive results) reduced with the use of DBT, implying greater reader confidence to classify a lesion as benign/normal or malignant and improving program accuracy by having:
 - fewer additional work-up views to determine malignancy or a benign structure
 - fewer cases of short interval follow-up (in programs/settings where this is the clinical response to an inconclusive mammogram result), reduced biopsy for women with a benign final outcome, and greater peace of mind and experience for women.
- BIRADS 1 or 2 lesion reporting (i.e., negative or benign results) increased meaning that more women did not have to undergo further imaging or follow-up (i.e., reduced workflow implications for assessment clinics), and/or

- BIRADS 4 and 5 lesion classifications (i.e., suspicious or highly suspicious results) increased in line with pathology-proven final malignant results (where available), suggesting improved predictive power for malignancy with DBT.

Increased reader confidence to dismiss or refer for biopsy is seen particularly in the reduction in indeterminate images (eg a change in the BIRADS 3 lesion classification).

Data from one prospective study and one large retrospective reader analysis in populations recalled to assessment looked at data pre- and post-implementation of DBT. They reported a decrease in the use of BIRADS 3 lesion reports and an increase in BIRADS 1 and 2 lesion reports with no decrease in sensitivity. This provides some assurance that mammographic findings can be more accurately classified when using DBT compared to than FFDM or DM alone. In these studies, fewer women were recalled to short-interval follow-up and were provided with a benign/normal finding with DBT. This is most likely due to improved lesion conspicuity and the availability of more information with which to assess margins that is available with DBT compared to FFDM. Given that BIRADS 3 lesions have a very low PPV for malignancy (typically less than 2%), increasing the accuracy of BIRADS 3 lesion classifications (without a commensurate decline in sensitivity) can improve diagnostic and health system performance. There is a commensurate reduction in unnecessary biopsy and fewer false positive results.

3.4.1. DBT results in a decrease in the number of inconclusive results compared to DM views (i.e., DBT results in the avoidance of more work-up for lesions with a benign final outcome)

Systematic reviews and narrative literature reviews

None of the systematic reviews reported on changes to final BIRADS lesion classification when an initial image was reassessed using DBT (either alone or as an adjunct to further work-up imaging).

One early narrative literature review (Tingberg & Zackrisson, 2011) reported on a small 2008 study by Andersson et al. (which preceded the inclusion date for *Allen + Clarke's* literature review). As described by Tingberg & Zackrisson, Andersson et al.'s study reported on 40 cancers with DBT images assessed by two experienced radiologists (no further details provided). They found that when using DBT images, readers provided higher BIRADS lesion scores for those classified as benign on DM. These lesions were then proven malignant at pathology. No further information about this study was presented in Tingberg & Zackrisson's study; however, later studies demonstrate consistency with these early results and with expectations of how DBT could perform (i.e., it reduces the effect of tissue overlap and increases lesion conspicuity therefore readers should have more information and be able to more accurately classify lesions as malignant or benign, resulting in more accurate triaging to further work-up or dismissal).

Prospective studies of symptomatic women only

The literature search returned four prospective studies that investigated DBT's influence on final BIRADS lesion classification. Chae et al. (2016) and Takamoto et al. (2013) recruited participants with pathologically-proven breast cancer or with lesions highly suspicious of cancers. Each study compared BIRADS lesion classification using DBT images (in one or two views) to FFDM images. Both studies reported similar findings: using DBT resulted in changes to BIRADS:

- BIRADS 3 lesion classifications reduced (resulting in fewer inconclusive results and fewer women undergoing short interval follow-up imaging)

- BIRADS 1 or 2 lesion classifications increased, and/or
- BIRADS 4 and 5 classifications increased in line with pathology results (where available).

In Chae et al.'s study, agreement between final BIRADS lesion classification and the reference standard (FNA cytology or histology of core needle or surgical biopsy or one-year follow-up for benign lesions) was 90.9% for DBT_{MLO} compared to 87.3% for FFDM ($p=.013$).

Similar changes to BIRADS lesion classification when using DBT were also reported by Takamoto et al (2013). In 99 Japanese women with pathologically-proven breast cancer, two radiologists rated 75.5% of affected breasts with the same BIRADS lesion classification overall; however, 23.5% were given a higher BIRADS score with DBT. This is an important study because Japanese women have more dense breast tissue, indicating that DBT may also result in more accurate diagnosis in women with dense breasts due to improved visibility of mass margin with DBT.

Prospective studies of mixed study populations (asymptomatic and symptomatic women)

Two studies (Bahrs et al., 2018; Mansour et al., 2014) included both women recalled to assessment and symptomatic women. Bahrs et al. (2018) investigated whether DBT alone affected the management of non-calcified lesions with no ultrasound correlate. Using 87 lesions (initially classified as BIRADS 3, inconclusive) with either histological results or FFDM follow-up imaging at six or 12 months, the authors reported statistically significant ($p<.0001$) findings that assessment with DBT (compared to FFDM):

- reduced overall BIRADS 3 lesion classifications (from 87 to 33 lesions) and therefore avoided further short interval follow-up for 54 women
- increased BIRADS 1 or 2 lesion classifications (a 51% increase), and
- decreased the proportion of focal asymmetries classified as BIRADS 3 lesions (-44.8%).

Four lesions were upgraded to BIRADS 4, two of which were T1c cancers that were detected on DBT before any follow-up imaging was conducted (i.e., the cancers were detected earlier). No false negatives or drops in sensitivity or specificity were reported in the two-year follow-up period. This led the authors to conclude that DBT accurately increased reader confidence to immediately dismiss normal and benign lesions and reduced recall to follow-up mammography for BIRADS 3 lesion classifications (including a reduction in additional radiation dose to women from avoided repeat interval screening).

Mansour et al. (2014) completed an analysis of 166 Egyptian women, comparing initial BIRADS lesion classification (FFDM) with a second-look assessment using FFDM + DBT. Like Bahrs et al., the authors concluded that FFDM + DBT (combo mode) enhanced the accuracy of BIRADS lesion classification. FFDM + DBT reduced overall BIRADS3 lesion classifications (from 52 to 15 lesions) and therefore avoided further work-up or short interval follow-up for 37 women, and increased BIRADS 1 or 2 lesion classifications (from 44 with FFDM to 70 with FFDM + DBT). Importantly, pathology-proven malignant cases were diagnosed much more accurately with FFDM + DBT (62.3%) compared to 18.1% with FFDM. Fewer cancers were missed with DBT (eight were missed on DBT compared to 38 missed with FFDM). The authors noted that the low value of data for FFDM was due to having a high proportion of women with dense breasts in the sample (69% had BIRADS 3 or 4 density). Significance testing was completed and reported as $p=.00$, but it is not clear exactly which data this related to.

Retrospective studies (general)

Most of the evidence regarding DBT's influence on BIRADS lesion classification is contained in retrospective reader studies: 16 studies presented results on initial and/or final BIRADS lesion classification agreement comparing FFDM/DM to DBT. Study designs generally involved reader review of a small set of images (often cancer-enriched and drawn from a larger cohort dataset) with comparisons made between either FFDM compared to FFDM + DBT, or FFDM/DM compared to DBT alone (in either one or two views). Although study design limitations limit the overall strength of the evidence, more recent research findings demonstrate effects that are broadly consistent with the prospective studies:

- use of DBT (either in combination with DM or alone) results in clinically significant reductions in BIRADS 3 lesion classifications (with consequently less additional supplementary imaging and fewer biopsies for a benign final outcome), and
- increases reader confidence in dismissing BIRADS categories 1 or 2 that may appear doubtful on FFDM, and increasing PPV₃ (i.e., increased BIRADS 5 lesion classifications with corresponding increase in biopsy-proven breast cancers).

All but one study reported this effect.

Retrospective studies of women recalled to assessment

FFDM + DBT compared to FFDM/DM (unspecified): BIRADS lesion classification agreement

Galati et al. (2017) completed a retrospective reader study of 214 symptomatic women, women with dense breasts and women with a previous breast cancer. All were participating in a screening program. For both readers, using FFDM + DBT significantly increased the number of lesions classified as BIRADS 1 or 2 and BIRADS 4 or 5 (both $p < .001$), and significantly reduced the number of recalls for further assessment (either additional imaging to complete the BIRADS assessment or short interval follow-up) ($p < .001$). As with other studies described in this section, Galati et al. concluded that FFDM + DBT in a clinical setting resulted in fewer additional work-ups compared to FFDM (based on increased reader confidence to dismiss normal/benign findings that were doubtful on FFDM).

DBT alone in one or two views compared to FFDM/DM (unspecified): BIRADS lesion classification agreement

In McDonald et al.'s (2017) study, six readers reviewed diagnostic work-ups for 168 women recalled from DM and 206 recalled from DBT to evaluate BIRADS 3 classification in a clinical setting. While the authors reported no statistically significant difference in the use of BIRADS 3 classification for any mammographic finding when FFDM or DBT was used, the use of DBT resulted in an overall decrease in the number of women recalled to assessment using DBT (2.4 fewer women per 1000 screening examinations who were recommended for follow-up).

Morel et al. (2014) reported data from a matched comparison reader study of 341 women either recalled to assessment or presenting symptomatically, finding that more lesions with benign or normal findings following triple assessment or histologically confirmed soft-tissue malignancies were correctly classified with 1v-DBT (not specified but presumably in same projection as the lesion was seen on initial imaging) compared to coned compression magnification mammography ($p = .0014$ and $p = .005$ respectively).

Brandt et al. (2013) reported changes in their retrospective reader evaluation of DBT as an alternative to DM work-up views. They reported a 50% decrease in BIRADS 3 classifications with DBT compared to DM (unspecified), a small increase in BIRADS 1 or 2 classifications, and almost no change in BIRADS 4 or 5 classifications.

Retrospective studies of only symptomatic women and/or women with a diagnosed breast cancer

Five studies of symptomatic women only compared final BIRADS classification.

FFDM + DBT compared to FFDM/DM (unspecified): BIRADS lesion classification agreement

Bansal & Young (2015) assessed 103 women undergoing triple assessment at a clinic for symptomatic women because of subtle mammographic findings or normal mammography but suspicious ultrasound (excluding calcified lesions and women with M5 classifications). The authors reported an 80% reduction in the use of M3 classifications when FFDM + DBT was used compared to FFDM. This result was consistent with but much higher than results reported in other studies using the combo imaging and was higher than results from studies comparing DM to DBT alone. In Bansal & Young's study, five biopsies were avoided (i.e., reductions seen in incorrect M4 classifications on FFDM). No statistical testing was reported but the authors concluded that FFDM + DBT imaging of women with subtle mammographic features is likely to result in more accurate lesion classification and reduced biopsy.

An important large case-control study by Raghu et al. (2016) compared 3576 women who received diagnostic DM imaging (not further specified) and three annual cohorts of women who were imaged with DBT following implementation into clinical practice. By year three, 99% (4676 women) of women having a diagnostic work-up were imaged with DBT. Comparing DM images and year 3 DBT images, Raghu et al. reported a statistically significant 50% decrease in the proportion of BIRADS 3 classifications, and an increase in BIRADS 1 and 2 classifications (from 58.7% for DM images to 75.8% for DBT images in Year 3, $p=.0001$). With DBT, fewer focal asymmetries were classified as BIRADS 3. More information about the radiological characteristics are discussed in *section 3.5*. No statistically significant changes between DM or DBT were made for BIRADS 4 or 5 classification but PPV₃ rose from 29.6% with FFDM to 50% with DBT in Year 3 ($p=.0001$).

Yang et al. (2013) included calcified lesions in their retrospective review of 59 pathologically proven breast lesions. The authors found that FFDM + DBT resulted in considerably fewer BIRADS 0 classifications (64 with FFDM compared to 10 with FFDM + DBT), with 78.2% of BIRADS 0 classifications for non-calcified lesions being upgraded to 4b or higher with FFDM + DBT. Further, BIRADS 5 classifications increased with FFDM + DBT (from 16 with FFDM to 48 with FFDM + DBT). The authors concluded that DBT as adjunct imaging resulted in more informative detection for all mammographic features of malignancy in women with more dense breasts, including microcalcifications. Further information about BIRADS lesion classification by primary mammographic feature from Yang et al.'s study is discussed in *section 4*. This study demonstrated that DBT provided more information to enable the radiologist to make a more definitive diagnosis without the need for further follow-up imaging.

Zuley et al. (2013) reported on the true positive and true negative fractions for soft-tissue breast lesions categorised as BIRADS 3 or higher (see *Table 9*), finding that, when using DBT to replace conventional mammographic diagnostic views (not further described), radiologists categorised fewer benign lesions as BIRADS 3 and that sensitivity remained high (false positive fraction was 0.85 with DM views and 0.74 with DBT). More lesions were classified as BIRADS 5 compared to FFDM.

While not reporting diagnostic accuracy using AUC measurement, Mercier et al. (2015) reported on sensitivity and specificity in a prospective study comparing DBT to DM, ultrasound and MRI. Using a cancer-enriched sample of 75 patients (all of whom had multifocal, multicentric, and

bilateral cancers resulting in 124 primary and satellite malignancies). The DM imaging work-up included additional or repeat DM/DSCV (eg, anterior, posterior, lateral oblique and additional – not specified), DBT, ultrasound and MRI. Using a scoring system based on the number of BIRADS > 3 lesions, tumour location histology, sensitivity, specificity and PPV/NPV, use of DBT resulted in more accurate lesion classification in women with breast density BIRADS 2 and 3 (4th edition) compared to DM work-up views. Sensitivity and PPV increased with the use of DBT compared to DM but specificity was lower for DBT with equal NPV (i.e., there was a higher rate of false positive classifications with DBT 74.4% compared to 81.6%). Lower (but not significant) specificity with DBT may have been due to reader bias as readers were aware of clinical presentation and staging of a BIRADS 4 or 5 lesion (and therefore the likelihood of cancer) and were therefore more likely to positively grade DM images.

Retrospective studies: diagnostic pathway unclear

FFDM + DBT compared to FFDM/DM (unspecified): BIRADS lesion classification agreement

A decrease in incomplete BIRADS assessment is similar to that reported by Waldherr et al. (2013). Reporting on their retrospective blinded reader study of 144 women recalled to assessment or with symptoms, Waldherr et al. (2013) found a statistically significant correct BIRADS classification with 1v-DBT (84%) and FFDM + DBT (85%) compared to FFDM (74.5%) ($p < .001$). Decreases in BIRADS 0 and increases in BIRADS 1, 2 and 5 categories were seen, and correct upgrading was recorded for 33/144 women and correct downgrading for 18/144 women, with the commensurate reduction in additional imaging for these women. That is, FFDM + DBT better classified benign and normal lesions compared to FFDM/DM and FFDM + DBT better supports accurate classification in lesions classified as indeterminate on FFDM/DM.

DBT alone in one or two views compared to FFDM/DM (unspecified): BIRADS lesion classification agreement

The remaining studies compared FFDM or DM (unspecified) to DBT alone (in one or two views), with similar results as those presented in the prospective studies

Improving sensitivity and reducing recall rate of non-calcified lesions was also a feature finding of Chan et al.'s 2017 fully crossed retrospective observer study, which was one of the only studies to use wide-angle DBT. The authors reported that with FFDM, 88 malignancies (mostly with subtle mammographic presentations) were classified as BIRADS <3 compared to 23 malignancies classified as BIRADS < 3 with DBT. A corresponding increase in BIRADS >4 classifications was seen with DBT compared to FFDM (no statistical testing reported).

Two small retrospective reader studies from Egyptian hospitals (Kamal et al., 2016; Taha Ali et al., 2016) both reported statistically significant increases in BIRADS 1 and 2 classifications, statistically significant decreases in BIRADS 3 classifications, and increases in BIRADS 5 classifications with DBT compared to FFDM. In Kamal et al.'s study, fewer lesions imaged with DBT were classified as BIRADS 3 (15.5% on DBT compared to 22.3% on FFDM; a down-grade of 12 cases) and 17 cases classified as BIRADS 4 on FFDM were also downgraded to BIRADS 1 or 2 when assessed using DBT images.

Lång et al. (2014) completed a side-by-side analysis of cases with discrepant BIRADS readings on DBT_{MLO} and FFDM, reporting that spiculated masses were more visible with DBT (see *section 4.5*) While not reporting directly on changes to BIRADS lesion classification between the two imaging modalities, Lång et al. (2014) completed a side-by-side analysis of cases with discrepant BIRADS classification on FFDM compared to DBT_{MLO/CC}. The main reasons for missed cancers with FFDM were reported to be due to the lesion being in dense breast tissue or therefore being

poorly visualised or due to subtle mammographic presentation. If the image was missed with DBT, it was more likely to be due to radiologist interpretive error (which could have been related to radiologist experience with DBT as an imaging modality). They reported that DBT improved lesion conspicuity both in terms of more views but also due the halo'ing effect of increased contrast between the lesion and background tissue. This was a specific issue for lesions with spiculated presentation (see *section 3.5.4* for further discussion).

Taken together and accepting that there is variation in the size of effect, all results from the retrospective studies add further evidence supporting improved radiologist performance in terms of improved accuracy in BIRADS lesion classification and consequent quicker but accurate ruling out of malignancy with DBT (which is most likely associated with reduced tissue overlap and better visualisation of breast architecture like radial scar or normal overlapping fibroglandular tissue with DBT compared to FFDM). These are important changes which should translate into fewer short-interval follow-ups and fewer biopsies in patients with benign lesions.

DBT compared to ultrasound: BIRADS agreement classification

Almost all the studies compared DBT to other mammographic views; however, Lee et al. (2016) compared diagnostic accuracy of ultrasound compared to DBT. Their retrospective study of 108 Asian women (69.7% of whom had dense breasts) with BIRADS 0 lesions (incomplete) reported fewer 3 and 4a classifications with DBT compared to ultrasound (14 compared to 21 BIRADS 3 classification) and more BIRADS 2 classifications. Three readers first read the screening mammogram and then the DBT image but were blinded to final pathology. It is not clear whether all three readers read the ultrasound images or whether just one reader did. This may have introduced bias into the study's results. They also reported that ultrasound underestimated the BIRADS classification in four malignancies (eg, the reader classified the lesions presenting as asymmetries with/without microcalcifications as 4b instead of 4c or 5, which was the classification reported following the viewing of the DBT image). This indicates that DBT showed better diagnostic performance and reduced unnecessary biopsies compared to ultrasound in women with more dense breasts (it is not clear whether these imaging work-up pathways were adjunctive to diagnostic mammography).

Table 9: Studies comparing BIRADS classification as a performance measure

9A Prospective studies

Study	Study details	Number of BIRADS 0 lesions	Number of BIRADS 1/2 or M 1/2 lesions	Number of BIRADS 3 or M3 lesions	Number of BIRADS 4 or M4 lesions	Number of BIRADS 5 or M5 lesions
Bahrs et al. (2018)	Prospective study in which participants had DBT not DSCV or magnification Participants: 87 consecutive patients (73 asymptomatic; 14 symptomatic) with non-calcified BIRADS 3 lesions with no ultrasound correlate Unit: Hologic Selenia Dimensions	No lesions graded as BIRADS 0	DBT Reader 1: 56 Reader 2: 56 Reader 3: 40	FFDM: 87 DBT: Reader 1: 27 Reader 2: 31 Reader 3: 42	DBT: Reader 1: 4 Reader 2: 4 Reader 3: 5	No lesions graded as BIRADS 5
Mansour et al. (2014)	Prospective reader (n=3) analysis blinded to pathology and other readers Participants: 166 consecutive Egyptian women (68.7% with dense breasts, BIRADS 3,4) recalled to assessment (n=72) or symptomatic (n=94), imaged with FFDM + DBT + ultrasound Unit: Hologic Selenia Dimensions	No data	FFDM: 44 FFDM + DBT: 70	FFDM: 47 FFDM + DBT: 15	FFDM: 56 FFDM + DBT: 28	FFDM: 14 FFDM + DBT: 48
Takamoto et al. (2013)	Prospective reader study (n=2) Participants: 99 Japanese women (195 breasts, 102 affected with breast cancer) with pathologically-proven breast cancer Unit: Hologic Selenia Dimensions	No data	<i>Affected breast only</i> FFDM: 18 DBT: 16 <i>15 of these cancers not seen on either FFDM or DBT</i>	<i>Affected breast only</i> FFDM: 26 DBT: 17	<i>Affected breast only</i> FFDM: 29 DBT: 28	<i>Affected breast only</i> FFDM: 29 DBT: 41

9B Retrospective studies comparing FFDM + DBT to FFDM

Study	Study details	Number of BIRADS 0 lesions	Number of BIRADS 1/2 or M1/2 lesions	Number of BIRADS 3 or M3 lesions	Number of BIRADS 4 or M4 lesions	Number of BIRADS 5 or M5 lesions
Galati et al. (2017)	Retrospective reader (n=2) study Participants: 214 women (37 cancer, 177 benign/normal); exclusions: BIRADS a, b; women with BRCA 1/2 mutations; pregnancy) Unit: Siemens Mammomat Inspirations	See BIRADS 3 (results combined, no further split described)	<i>Reader 1</i> FFDM: 63 DBT: 129 <i>Reader 2</i> FFDM: 47 DBT: 136	<i>Reader 1</i> FFDM: 112 DBT: 17 <i>Reader 2</i> FFDM: 131 DBT: 17	<i>Reader 1</i> FFDM: 39 DBT: 68 <i>Reader 2</i> FFDM: 36 DBT: 61	No data
Bansal & Young (2015)	Retrospective reader (n=4) analysis with blinding to final histology and patient history Participants: 103 consecutive Welsh women with subtle mammographic or mammographically normal by suspicious ultrasound (16 malignancies; 90 normal); key excluded patients: BIRADS M5; microcalcifications Unit: Hologic Selenia Dimensions	No data	FFDM: 7 images FFDM + DBT: 75	FFDM: 91 images FFDM + DBT: 18 images	M4 and M5 FFDM: 8 images FFDM + DBT: 13 images	M\$ and M% combined - see previous column
Waldherr et al. (2013)	Retrospective blinded reader (n=2) study Participants: 144 women either recalled to assessment or with symptoms Unit: Hologic Selenia Dimensions prototype	FFDM: 53 DBT _{MLO} : 19 FFDM + DBT: 15	FFDM: 7 DBT _{MLO} : 35 FFDM + DBT: 31	FFDM: 7 DBT _{MLO} : 2 FFDM + DBT: 5	FFDM: 64 DBT _{MLO} : 53 FFDM + DBT: 66	FFDM: 13 DBT _{MLO} : 35 FFDM + DBT: 27
Yang et al. (2013)	Retrospective, multi-reader (n=3) study Participants: 171 radiologist scores for 59 pathologically proven breast cancers in Taiwanese women (17 masses, 12 focal asymmetry/density, 6 AD, 23 microcalcifications, 1 other) Unit: Hologic Selenia Dimensions	FFDM: 64 FFDM + DBT: 10	FFDM: 3 FFDM + DBT: 2	FFDM: 2 FFDM + DBT: 0	<i>All BIRADS 4</i> FFDM: 86 FFDM + DBT: 111 <i>BIRADS 4a</i> FFDM: 33 FFDM + DBT: 29 <i>BIRADS 4b</i> FFDM: 33 FFDM + DBT: 40 <i>BIRADS 4c</i> FFDM: 20 FFDM + DBT: 42	FFDM: 16 FFDM + DBT: 48

9C Retrospective studies comparing DBT alone (in one or two views) to FFDM/DM (unspecified)

Study	Study details	Number of BIRADS 0 lesions	Number of BIRADS 1/2 or M 1/2 lesions	Number of BIRADS 3 or M3 lesions	Number of BIRADS 4 or M4 lesions	Number of BIRADS 5 or M5 lesions
Chan et al. (2017)	Fully crossed, retrospective reader (n=6) study comparing DBT to FFDM in enriched dataset Participants: 134 cases recalled to assessment or with clinical findings, excluding microcalcification cases (61 malignancies, 85 benign/normal) Unit: GE GEN2 prototype (wide angle 60°)	No data (used forced BIRADS methodology)	<i>BIRADS 3 and below</i> FFDM malignancies: 88 (22 AD, 26 asymmetries, 30 masses) DBT malignancies: 23 (1 AD, 8 asymmetries, 23 masses) FFDM benign lesions: 280 DBT benign lesions: 266	BIRADS 3 and below used – see previous column for results	<i>BIRADS 4a and above</i> FFDM malignancies: 278 (56 AD, 66 ASD, 156 masses) DBT malignancies: 343 (77 AD, 94 ASD, 172 masses) FFDM benign lesions: 206 DBT benign lesions: 220	BIRADS 4a and above used – see previous column
McDonald et al. (2017)	Retrospective reader (n=6) audit of women with BIRADS 3 at diagnostic mammogram Participants: FFDM: 1112 women and DBT: 1366 women with similar baseline characteristics Unit: Hologic Selenia Dimensions	No data	No data	FFDM: 38.6% DBT: 40.5%, $p=.69$ <i>BIRADS 3 category per 1000 screening examinations:</i> FFDM: 114.5 DBT: 119.8, $p=.721$	No data	No data
Kamal et al. (2016)	Retrospective reader (n=2) analysis Participants: 98 symptomatic Egyptian women (103 lesions; 78 benign, 25 malignant) and more dense breasts (BIRADS 3 or 4) Unit: Hologic Selenia Dimensions	No data	FFDM: 20.4% DBT: 47.6% MRM: 54.4%	FFDM: 22.3% DBT: 15.5% MRM: 16.5%	FFDM: 57.3% DBT: 21.4% MRM: 15.5%	FFDM: No data DBT: 15.5% MRM: 13.6%
Raghu et al. (2016)	Retrospective case-control study Participants: 3576 FFDM images (Y0); - 3394 DBT images (Y1); 4541 DBT images (Y2); 4676 DBT images (Y3) Unit: Hologic Dimensions	No data	<i>DM</i> Y0: 58.7% <i>DBT</i> Y1: 67.3% Y2: 69.8% Y3: 75.8%	<i>DM</i> Y0: 33.3% <i>DBT</i> Y1: 25.1% Y2: 21.8% Y3: 16.4%	<i>DM</i> Y0: 8.0% <i>DBT</i> Y1: 7.6% Y2: 7.9% Y3: 7.8%	See previous column (BIRADS 4 and 5 combined)

Study	Study details	Number of BIRADS 0 lesions	Number of BIRADS 1/2 or M 1/2 lesions	Number of BIRADS 3 or M3 lesions	Number of BIRADS 4 or M4 lesions	Number of BIRADS 5 or M5 lesions
Taha Ali et al. (2016)	Retrospective reader (n=3) study at a single institution Participants: 132 Egyptian women with at least one lesion (67 malignancies: 34 IDC, 17 ILC, 7 IDC + DCIS; 9 others; 78 benign) detected with DM +/- ultrasound Unit: GE SenoClaire	No data	FFDM: 39 DBT: 62	FFDM: 41 DBT: 14	FFDM: 49 DBT: 25	FFDM: 12 DBT: 40
Haq et al. (2015)	Retrospective sub-study Participants: 31 malignant abnormalities Unit: Hologic Selenia Dimensions <i>TOMMY trial sub-study</i>	No data	DM _{MLO} : 1 DBT _{MLO} : 0 DM _{CC} : 4 DBT _{CC} : 0	DM _{MLO} : 6 DBT _{MLO} : 1 DM _{CC} : 14 DBT _{CC} : 6	DM _{MLO} : 0 DBT _{MLO} : 2 DM _{CC} : 8 DBT _{CC} : 2	DM _{MLO} : 6 DBT _{MLO} : 9 DM _{CC} : 4 DBT _{CC} : 12
Morel et al. (2014)	Retrospective matched reader (n=7) study comparing diagnostic accuracy of FFDM + 1v-DBT (not specified) to FFDM + CCMM Participants: 341 women (103/354 lesions were malignant and of these 80 were soft-tissue and 23 were DCIS) recalled for assessment following screening or symptomatic mammography Unit: Hologic Selenia Dimensions	No data	<i>Normal/benign lesions</i> FFDM + CCMM: 135 FFDM + 1v-DBT: 162 <i>All malignancies</i> FFDM + CCMM: 5 FFDM + 1v-DBT: 1 <i>Soft tissue malignancies</i> FFDM + CCMM: 4 FFDM + 1v-DBT: 0	<i>Normal/benign lesions</i> FFDM + CCMM: 90 FFDM + 1v-DBT: 72 <i>All malignancies</i> FFDM + CCMM: 25 FFDM + 1v-DBT: 18 <i>Soft tissue malignancies</i> FFDM + CCMM: 15 FFDM + 1v-DBT: 7	<i>Normal/benign lesions</i> FFDM + CCMM: 22 FFDM + 1v-DBT: 9 <i>All malignancies</i> FFDM + CCMM: 30 FFDM + 1v-DBT: 30 <i>Soft tissue malignancies</i> FFDM + CCMM: 17 FFDM + 1v-DBT: 16	<i>Normal/benign lesions</i> FFDM + CCMM: 4 FFDM + 1v-DBT: 8 <i>All malignancies</i> FFDM + CCMM: 43 FFDM + 1v-DBT: 54 <i>Soft tissue malignancies</i> FFDM + CCMM: 38 FFDM + 1v-DBT: 51
Brant et al. (2013)	Retrospective, blinded reader (n=3) evaluation of DBT as an alternative to conventional DM views (range 1-6 views) in the workup of noncalcified asymmetries, areas of distortion, and masses Participants: 146 women recalled to assessment mammography with 158 abnormalities Unit: Hologic Selenia Dimensions beta unit	No data	<i>With DBT</i> Reader 1: 137 Reader 2: 137 Reader 3: 129 <i>With diagnostic mammography</i> Reader 1: 134 Reader 2: 134 Reader 3: 134	<i>With DBT</i> Reader 1: 4 Reader 2: 3 Reader 3: 4 <i>With diagnostic mammography</i> Reader 1: 8 Reader 2: 8 Reader 3: 8	<i>With DBT</i> Reader 1: 16 Reader 2: 16 Reader 3: 14 <i>With diagnostic mammography</i> Reader 1: 16 Reader 2: 16 Reader 3: 16	See previous column (BIRADS 4 and 5 combined)

3.5. Radiation dose

In the BSA program, up to 11.8% of women undergoing a prevalent breast screen and 3.8% percent of women completing incident screening will be recalled for further assessment due of a potentially suspicious finding.¹³ If recalled for assessment, a woman is likely to undergo supplemental DM views (in centres where DBT is not used). Additional work-up views are designed to enable closer consideration of a suspicious lesion, confirm the presence of a breast cancer and gather further information to support treatment planning. Additional DM views can improve spatial resolution, solve problems associated with the superimposition of overlapping tissue and acquire exact coordinates to enable accurate targeting for biopsy (if required). With each additional DM view, the total dose of radiation a woman is exposed to increases, with some women having more than three additional views after an initial FFDM screening examination.

Both DBT and DM are radiation-emitting procedures. Dose is cumulative. Considering the radiation dose of different imaging protocols is an important safety consideration for any breast care team and breast-screening program. A review of the literature comparing the effective radiation dose between DBT and FFDM is available in a previous *Allen + Clarke* literature review.

Findings from *Allen + Clarke's* literature review on the role of DBT in screening

Radiation dose varies with the image acquisition process used (DBT or FFDM or combination mode), the number of and type of views, the use of automatic exposure control, breast size and composition, compression needed to acquire the image, patient positioning, time needed to acquire the image, and by DBT unit used.

Almost all the studies included assessed MGD for a screening strategy based on FFDM + DBT compared to FFDM (i.e., the combination mode) and most were completed using Hologic's Selenia Dimensions unit. In most studies, the MGD was doubled with FFDM + DBT compared to FFDM but was still below maximum per view limits. Other imaging protocols were also reported as researchers investigated ways to maintain the benefits of DBT with a reduction in MGD. Other reported results were that DBT alone (in two views) resulted in a similar radiation dose compared with FFDM; that DBT_{MLO} resulted in a lower radiation dose compared to FFDM; and that FFDM + DBT_{MLO} resulted in a slightly higher dose compared to FFDM.

More recent studies have investigated the efficacy of DBT + s2DM, which eliminates the need for a separate 2D image acquisition. Using this approach, 2D DM images are synthesised from a 3D DBT-acquired dataset. This approach halves the effective dose of combined FFDM + DBT, making dose comparable to FFDM.

Moving to FFDM + DBT as the preferred screening strategy could have significant implications for cumulative dose if separate acquisitions are used for 2D and 3D images, if the screening interval is annual rather than biennial, or if women start participating in mammography-based breast cancer screening in their early 40s.

As reported in *section 3.1* to *section 3.5* of this literature review, using DBT provides more accurate imaging which allows for better lesion identification and classification; however, a major hinderance for widespread adoption appears to be the increased radiation dose to patients. FFDM + DBT is the imaging protocol for which we have the most complete and consistent evidence of improved diagnostic accuracy; however, this is also the imaging protocol that appears to expose women to the highest radiation dose. Reducing MGD while optimising

¹³ From <https://ncci.canceraustralia.gov.au/screening/abnormal-breast-screen-assessment/recall-assessment>, Accessed 5 July 2018.

diagnostic performance of breast imaging with DBT is an important clinical and research focus. Some clinics and researchers modify the number of view combinations in which they use DBT, whereas others have developed algorithms that create synthetic 2D images from the array of DBT slices as a means of replacing FFDM.

To extend the previous evidence reported in *Allen + Clarke's* literature review on DBT in screening, our literature search focused on identifying papers that assessed the total radiation dose of additional follow-up DM images required if a woman is recalled to assessment (i.e., MGD for FFDM + DSCV compared to FFDM + DBT of FFDM compared to s2DM). Papers discussed in the literature review on tomosynthesis in screening were excluded.

A total of 15 studies and three narrative reviews reported on radiation dose. Four studies compared the MGD for FFDM + DSCV compared to DBT; two studies compared MGD for FFDM to s2DM; and the remaining studies reported MGD only. The studies discussed in this literature review are listed below.

Systematic and/or literature reviews

Systematic reviews: none

Three narrative literature reviews: Eghtedari et al. (2018); Destounis (2017); Gartner Roth (2014)

RCTs and prospective studies

Four studies:

Only women recalled to assessment: Bahrs et al. (2018); Caumo et al. (2018); Tagliafico et al. (2012)

Diagnostic pathway unclear: Alakhras et al. (2014);

Retrospective studies

Eleven studies:

Only women recalled to assessment: Gilbert et al. (2015b); Morel et al. (2014); Brandt et al. (2013)

Women recalled to assessment and symptomatic women: Kang et al. (2016); Wallis et al. (2012)

Only symptomatic women and/or women with a diagnosed breast cancer: Kim (2016); Seo et al. (2016); Bansal & Young (2015)

Diagnostic pathway unclear: Endo et al. (2017); Choi et al. (2016b); Elizalde et al. (2016)

Key findings

Per view, the MGD is higher for FFDM + DBT compared to FFDM alone or additional spot views

Overall reported real-world MGD for DBT compared to FFDM views were similar to that reported in *Allen + Clarke's* previous literature review: per view, DBT alone has a similar MGD compared to DM but the combined mode (FFDM + DBT) exposes women to almost double the radiation dose. Therefore, investigating ways to reduce the dual acquisition dose remains an important area of study.

DBT's MGD per view should be considered in the context of lifetime radiation dose from all mammography (screening + additional assessment imaging)

Intuitively, there is likely to be considerable variation in the average difference between MGD for an imaging work-up involving FFDM + DBT compared to FFDM + DSCV because women are likely to require a different number of DSCV views to collect all the necessary information to inform diagnosis and treatment planning. Less variation in dose would be expected with DBT as the necessary information is likely to be available in one or two views. Further research is needed to confirm this. While the DBT imaging dose is higher than FFDM, women are less likely to be recalled for further assessment and therefore experience both fewer instances of supplementary DSCV and, if DBT provides sufficient diagnostic information, avoided overall DM work-up across her lifetime.

Efforts to optimise diagnostic accuracy and reduce radiation dose continue

Research to determine the lowest possible radiation dose needed to acquire satisfactory images continues. There is some emerging evidence that DM_{CC} + DBT_{MLO} results in minimal increase in MGD with a good improvement in lesion detection; however, further work to test the diagnostic accuracy of this imaging protocol is required, including considering how it might interface with s2DM, which also results in decreased radiation dose but performance that is equivalent to FFDM + DBT.

3.5.1. Per view, the MGD is higher for FFDM + DBT compared to FFDM alone or additional spot views

Systematic reviews and narrative literature reviews

None of the systematic reviews provided detailed commentary on the comparative MGD of FFDM + DSCV compared to DBT views.

Destounis' (2017) narrative review discussed that, while the MGD for FFDM + DBT does not exceed the ACR Mammography Quality and Standards Act (MQSA) dose limit of 3.0 mGy per breast per view, it does significantly increase the total amount of radiation that women are exposed to. Reported MGD for:

- FFDM was 3.77 mGy, and
- FFDM + DBT was 7.97 mGy.

Prospective studies

Bahrs et al. (2018) evaluated the influence of DBT on the management of 105 consecutive women with non-calcified BIRADS 3 findings with no obvious ultrasound correlate (i.e., inconclusive screening results). All patients were imaged with DBT (Hologic Selenia Dimensions) instead of DSCV. The MGD was collected with each image, with DBT having a higher MGD compared to FFDM:

- FFDM: MGD for each breast was 3.1 mGy (range: 1.8-8.67 mGy), and
- DBT: MGD for each breast was 4.9 mGy (range: 2.29-9.6 mGy).

Other studies reported on MGD; however, they did not provide comparative analysis between the DBT imaging protocol and FFDM + DSCV. Reported MGD are similar to the results presented in Allen + Clarke's literature on DBT in screening: MGD for DBT alone is comparable to FFDM and approximately doubled with used as FFDM + DBT. A summary of results is presented in *Table 10 (A-B)* (overleaf).

Table 10: Reported MGD

10A Prospective studies

Study	Sample	DM MGD (mGy)	DBT MGD (mGy)	DM + DBT MGD (mGy)
Caumo et al. (2018)	Prospective pilot evaluation of 34,017 Italian women (315 malignancies) participating in a population-based screening program who had DBT Unit: Hologic Dimensions unit	<i>Single DM view</i> 1.48 ± 0.58 (SD: 0.52, 3.13)	<i>Single DBT view</i> 2.09 ± 0.55 (SD: 1.13, 3.65)	No data
Alakhras et al. (2014)	Prospective reader (n=26) performance study of 50 cases (32 malignancies) Unit: Hologic Selenia Dimensions	1.9	2.1	No data
Wallis et al. (2012)	Prospective; multi-centre study of women with abnormal images (40 malignancies) Unit: MicroDose Mammography system	English: 1.2 Swedish: 0.6	English: 0.82 Swedish: 0.7	No data

10B Retrospective studies

Study	Sample	DM MGD (mGy)	DBT MGD (mGy)	DM + DBT MGD (mGy)
Elizalde et al. (2016)	Retrospective reader study of 1041 Spanish women and 1 man with biopsy-proven malignancies (n=84) Unit: Siemens Inspiration	1.9	1.9	3.8
Kim et al. (2016)	Retrospective comparison analysis using a prospective cohort of 172 Korean women with 184 cancers Unit: Hologic Selenia Dimensions	1.2	1.3	No data
Seo et al. (2016)	Retrospective reader (n=6) evaluation (data collected prospectively) of 203 symptomatic women or recalled from mammography or ultrasound (129 malignancies) Unit: Hologic Selenia Dimensions	CC: 1.68 ± 0.68 MLO: 1.62 ± 0.75	CC: 1.17 ± 0.51 MLO: 1.65 ± 0.61	No data
Gilbert et al. (2015b)	Retrospective, multicentre (n=6), blinded multi-reader (n=26) matched study (TOMMY trial) of 7060 women recalled to assessment or aged <50y with a family history of breast cancer (1137 malignancies) Unit: Hologic Selenia Dimensions and C-view software	3 mGy	4 mGy	7 mGy
Morel et al. (2014)	Retrospective matched reader (n=7) study of 341 women (103 malignancies) Unit: Hologic Selenia Dimensions	1.2	1.45	No data

3.5.2. DBT's MGD per view should be considered in the context of lifetime radiation dose from all mammography (screening + additional assessment imaging)

Systematic reviews and narrative literature reviews

Gartner Roth et al. (2014) reported that overall radiation dose with DBT imaging is likely to result lower lifetime radiation dose because, while the DBT imaging dose is higher than FFDM, women are less likely to be recalled for further assessment and therefore experience both fewer instances of supplementary DSCV and, if DBT provides sufficient diagnostic information, avoided overall DM work-up. This concept was also explored by Bansal & Young (2015) and Brandt et al. (2013).

Prospective studies

Bahrs et al. (2018) authors found that DBT resulted in a significantly higher dose (1.5x higher, $p < .001$) compared to FFDM; however, they noted that the radiation dose of DBT was lower than FFDM + DSCV (although mGy values were not provided); and theorised that the use of DBT could reduce the total radiation dose patients receive.

Another early small study (52 women) reported that the replacement of DSCV with DBT could lower MGD. Tagliafico et al. (2012) reported that radiation dose with DBT was lower than the combination of FFDM + DSCV: 2.39 ± 0.6 mGy compared to 4.69 ± 1.7 mGy.

Retrospective studies

In an early retrospective study of women recalled to assessment, Brandt et al. (2013) evaluated DBT as an alternative to conventional diagnostic DM views in the assessment of noncalcified asymmetries, distortions, and masses. Brandt et al. did not report total radiation dose but reported that DBT could reduce the number of additional DM views obtained during a diagnostic workup because it does not require the same degree of positioning accuracy to obtain clear images. Since additional diagnostic DM imaging can include MLO, DSCV, and/or magnification views, and/or rolled or tangential views, the total MGD could easily add up to a similar or greater dose than FFDM + DBT if used as a primary screening test, or FFDM as a screening test followed by DBT for the further assessment of suspicious lesions.

Considering additional work-up images in symptomatic women, Bansal & Young reported a figure of 2.5 mSv for FFDM + DSCV. In addition, there may be variation in the average difference between MGD with FFDM + DBT compared to FFDM + DSCV as women are likely to require a different number of DSCV views to collect all the necessary information to inform diagnosis and treatment planning. Less variation would be expected with DBT as the necessary information is likely to be available in one or two views, again contributing to a lower overall lifetime radiation dose.

3.5.3. There are ways to optimise image quality while lowering MGD, especially though the use of synthesised views

Systematic reviews and narrative literature reviews

Eghtedari et al. (2018) reported on a study by Shin et al. in which the authors explored maximum cancer detection coupled with minimal increase in radiation dose. Shin et al. reported that $DBT_{MLO} + DM_{CC}$ resulted in the best imaging and minimal increase in exposure when compared to FFDM. Interestingly, this imaging protocol was also used by Kang et al. (discussed below).

Destounis (2017) also reported on studies showing that the use of s2DM resulted in reductions in MGD of approximately 39% (as reported in Zimmerman et al., 2012). More recently, as cited

by Destounis, in a scientific exhibit poster presentation, Garayoa et al. (2014) reported that replacing FFDM + DBT with DBT + s2DM could reduce the MGD by 43%. This figure was again repeated in Garayoa et al.'s 2018 article; however, no updated specific dose data was reported.

Retrospective studies

Two studies reported on the MGD of different imaging modalities.

In the study by Kang et al. (2016), the diagnostic performance of s2DM (using GE SenoClaire with Volume preview software) was compared against one view of DM to determine a means of reducing radiation dose. The authors found the combined use of DM_{CC} + DBT_{MLO} + s2SM showed higher sensitivity and specificity to FFDM with only a slightly higher overall MGD (9.3%, 5.78 mGy \pm 1.07). Specific results were:

- FFDM: 5.29 mGy, compared to
- DM_{CC} + DBT_{MLO} + s2DM: 8.44 mGy.

These results reflect the reported optimal balance between detection and dose as discussed by Eghtedari et al. (2018). Kang et al.'s study was a single-centre, retrospective study but the results show promise for a sound combination to bring the benefits of the DM and DBT together without substantially increasing the radiation dose.

The other study reporting on radiation dose and s2DM compared to FFDM was Choi et al. (2016b). Choi et al. showed that the diagnostic performances of s2DM and FFDM were equivalent in detecting T1-stage breast cancers suggesting that conventional FFDM with DBT imaging may be unwarranted. Without the need for standard FFDM imaging this eliminates "the double-dose" radiation effect. The difference in reported MGD between s2DM and FFDM for a single view was almost indistinguishable with a difference of 0.03 mGy (1.73 mGy for FFDM and 1.70 mGy for DBT)

These results provide some evidence that if conventional DM imaging is used more sparingly (or is replaced by DBT), radiation safety becomes less of an issue.

Endo et al. (2017) investigated a new method of creating a similar quality image to a conventional DBT image with lower radiation and still retaining the same number of views. The MGD for a single view was higher with the DBT combination:

- 1.59 \pm 0.29 mGy for DBT (nd) + FFDM (nd)
- 1.02 \pm 0.21 mGy for FFDM (nd) alone, and
- 1.70 \pm 0.49 mGy for conventional FFDM.

The difference in MGD between DBT (nd) + FFDM (nd) and conventional FFDM was – 0.11 mGy and was not significant (95% CI: – 0.147, 0.008; P = 0.081). While the specificity of DBT (nd) + FFDM (nd) was superior to conventional FFDM, this was the first study investigating this technology and took place in a single institution. Research into a wider scope of clinical practice is needed before any recommendations can be made. Additionally, this method should be compared to s2DM to identify which is the more effective technology for reducing MGD without compromising their ability to detect cancers.

4. MAMMOGRAPHIC FINDINGS: IMPROVING CONSPICUITY

Suspicious mammographic findings may include distortion of normal breast tissue without a corresponding mass, a difference in breast density between the breasts (a global or focal asymmetry), specific patterns of microcalcifications and masses. Some primary mammographic features are clearly seen on DM (such as microcalcifications and some masses), whereas other presentations (like AD or asymmetries) can be subtle and much more difficult to detect.

Subtle differences in x-ray attenuation makes differentiating between malignant and benign breast tissues in the assessment of breast cancer a complex process. Using methods that improve visibility and clarity of primary mammographic findings (especially subtle findings like AD or focal asymmetry) has a positive impact on readers' ability to detect areas suspicious for cancer and can improve diagnostic accuracy and reader performance. Almost all the studies reported in this literature review explain, in brief, how DBT addresses some of the issues associated with tissue superimposition seen when using DM. This superimposition can mean that normal breast structures mimic suspicious lesions or that malignancies can be hidden by overlapping tissue (especially in women with more dense breasts). Additional imaging may be needed to solve issues that can be attributed to overlapping tissue (rather than a malignancy). With DBT, tissue overlap is reduced, which in turn increases visibility of lesion margins and overall conspicuity and makes it easier to distinguish between normal or benign breast structures and breast cancers. Overall improved conspicuity with DBT is well-accepted in the literature. Further specific considerations relating to lesion conspicuity include that:

- curved or circumscribed margins (most likely indicative of a benign lesion) are more clearly seen on DBT, making it easier for readers to confidently dismiss benign masses and eliminate unnecessary further imaging work-up or biopsy (Michell & Batohi, 2018; Moseley, 2016), and
- AD and spiculated or stellate lesions (i.e., those which might present as a star-shape or spiky) are generally better seen with DBT compared to DM (Houssami et al., 2016).¹⁴

Investigating the impact of improved visibility is an important dimension when considering DBT's place in a breast cancer assessment clinic. Thirty-three studies and nine narrative literature reviews investigated differences in DM performance compared to DBT as it related to a primary mammographic finding. This section describes evidence comparing FFDM to DBT for the following primary mammographic findings:

- AD
- asymmetry
- microcalcifications, and
- masses.

4.1. Compared to FFDM, DBT improves lesion conspicuity, making it easier for readers to determine whether a lesion is suspicious for malignancy

Improved overall conspicuity has been demonstrated in several early studies in which difference in conspicuity by imaging technique was reported. Houssami et al. (2016) reported that DBT

¹⁴ Houssami et al. also noted that DBT findings from the Oslo Tomosynthesis in Screening, Malmö and STORM trials increased invasive cancer detection and that these had mammographic findings of spiculated mass or AD.

provides better imaging of very small or subtle mammographic findings such as spiculated masses, AD with or without spiculation, and stellate masses and distortions.

In studies based in screening populations or women recalled to assessment:

- Nam et al. (2015) investigated mammographically occult cancers (i.e., those detected with ultrasound not FFDM) in 41 women with predominantly dense breasts. They reported that when an unblinded review of the DBT and FFDM images was undertaken (i.e., the readers were told where the lesions were), 25 were visible on FFDM compared to 34 that were visible on DBT ($p=.047$). A rating system was used to classify visibility and with this, 22 lesions were constantly visible on DBT but only 11 were constantly visible with FFDM. The study had three readers with less than 10 years' experience and the authors reported moderate inter-reader agreement (ICC = 0.504 and 0.679). Lesion visibility was also improved with 54% of cancers detected with ultrasound (and mammographically occult) were also detected with DBT compared to 29% which were detected with DBT when the suspicious area was known.
- Skaane et al. (2012) reported that three readers with more than five years' experience as breast radiologists rated the conspicuity of 23 cancers in a side-by-side feature analysis. Conspicuity was only greater with FFDM for three cancers presenting as a mass or density (i.e., 20 cancers had equivalent or better conspicuity with DBT).
- Poplack et al. and Good et al. (cited in Tingberg & Zackrisson, 2011) reported that lesion conspicuity with DBT was equivalent to or better than FFDM in 89% and 98% of cases respectively.
- Tagliafico et al. (2012) reported that DBT provided superior conspicuity compared to DSCV, with statistically significant mean conspicuity values of 4.1 ± 0.3 for DBT and 2.9 ± 0.4 for FFDM, $p < .001$. Improved lesion conspicuity is likely to improve diagnostic accuracy.

Studies in diagnostic populations (or populations of symptomatic women) report similar findings:

- Lång et al. (2014) completed a side-by-side analysis of cases with discrepant BIRADS reporting outcome on DBT compared to DM_{MLO} or CC. The authors noted that DBT's better visualisation of lesions (specifically spiculated tumours) created an advantage over DM. Additionally, Lång et al. found evidence to suggest that reasons for missing tumours differed between modalities. Missing a tumour observed on DM was frequently due to readers being unable to clearly see the lesion (tissue overlap and lower conspicuity of radiographic features). On DBT, the most common reason for missing a lesion was due to interpretive error regarding clearly visible lesions. Therefore, this error may diminish with increased reader experience.
- In Yang et al.'s (2013) study of women who had pathologically proven breast cancer, FFDM + DBT was compared to FFDM. Three readers perceived that FFDM + DBT was more informative in 48% of all cases and was considerably better in the depiction for 83% of masses and 94.4% of AD presentations. The overall rating is pulled down by the microcalcification results (which was 11.6% of calcified lesions were better depicted in DBT).

- Hakim et al. (2010) used four radiologists to complete an early, small side-by-side pilot review of 25 FFDM and DBT images from both asymptomatic and symptomatic women to determine if this DBT images were seen better, worse or comparable to FFDM and other mammographic views. Radiologists' thought:
 - FFDM + DBT views were better for diagnosis in half of the cases and better visibility was reported for all eight high-risk lesions (including atypical ductal hyperplasia and intraductal papilloma)
 - at least equivalent in one-third of cases, and
 - poorer in less than one-fifth of cases (in cases where subjective rating was negative for FFDM + DBT, breast density was either scattered or heterogeneously dense, or because the lesions seen better in the DSCV image were optimally positioned and where they were not seen on the DBT imaging, and the radiologist would have dismissed the lesion as benign without the need for ultrasound).

4.2. Architectural distortion

The BIRADS classification system defines AD as being when:

'the normal architecture of the breast is distorted with no definite mass visible. This includes spiculations radiating from a point and focal retraction or distortion at the edge of the parenchyma' (cited in Gaur et al., 2013).

Breast cancers presenting with a subtle primary mammographic finding like AD can be difficult to perceive on DM. Detection of AD has relatively low inter-reader agreement (for example, Mariscotti et al.'s ROC curve analysis for AD presentation was low for both DBT and FFDM and was less than kappa statistic 0.7), indicating both the difficulty readers may have in detecting subtle AD, and the ease at which a subtle AD presentation can be missed, misinterpreted or misclassified. AD is a common finding on review of false-negative mammograms: Destounis (2018), Peppard et al. (2015) and Partyka et al. (2014) all reported evidence suggesting that 12-45% of cancers missed on screening mammography were detectable via subtle AD presentation. Further, AD may indicate invasive breast cancer (particularly ILC and IDC) and is the primary mammographic finding in approximately four percent of ductal carcinoma in situ (DCIS) cases (Gaur et al., 2013). Correct identification of AD may result in earlier detection of breast cancer (and therefore a better prognosis for patients). Complicating assessment is the fact that AD may also indicate benign structures (like radial scar, sclerosing adenosis, fat necrosis, complex sclerosing lesions or post-procedural changes).

This literature review describes findings from one systematic review, 17 primary studies and five narrative literature reviews that reported on detection of AD with DBT (either as a primary study outcome or within a wider radiologic feature analysis). As AD is a subtle mammographic finding, study samples often contain only a small number of breast cancers presenting with AD as the primary mammographic feature. Most of the relevant data is presented in retrospective studies. This limits the strength of the evidence base relating to the detection of AD with DBT compared to other imaging techniques. The studies discussed in this literature review are listed below. A summary of findings is included in *Table 11 (A-H)*.

Systematic reviews and narrative literature reviews

One systematic review: Yun et al. (2017)

Four narrative literature reviews: Michell & Batohi (2018); Destounis (2017); Mall et al. (2017); Peppard et al. (2015)

One pictorial atlas: Houssami et al. (2016)

RCTs and prospective studies

RCTs: none

Four prospective studies:

Only women recalled to assessment: Caumo et al. (2018)

Women recalled to assessment and symptomatic women: Skaane et al. (2012)

Only symptomatic women and/or women with a diagnosed breast cancer: Refaat & Matar (2015)

Diagnostic pathway unclear: Bahrs et al. (2017)

Retrospective studies

Thirteen retrospective studies:

Only women recalled to assessment: Dibble et al. (2018); McDonald et al. (2017); Ni Mhuircheartaigh et al. (2017); Gilbert et al. (2015a); Partyka et al. (2014)

Women presenting for screening and symptomatic women: Taskin et al. (2017); Ray et al. (2015)

Only symptomatic women and/or women with a diagnosed breast cancer: Mariscotti et al. (2016); Raghu et al. (2016); Takamoto et al. (2013); Yang et al. (2013)

Diagnostic pathway unclear: Chan et al. (2017); Mariscotti et al. (2017); Freer et al. (2015)

Key findings

AD presentation is a subtle mammographic presentation that readers may have trouble detecting on DM. It can easily be missed, misinterpreted or misclassified. AD is a common finding on review of false-negative mammograms. It may indicate invasive breast cancer (particularly ILC and IDC). It can also indicate a benign finding. Correct identification of a benign lesion or malignancy is important so that cancers are not missed and that women do not undergo unnecessary assessment or diagnostic work-up.

Detection of AD, which is often a subtle mammographic presentation for invasive breast cancer, increases with the use of DBT because of improved visibility

Compared to FFDM, DBT improves AD conspicuity and increases the overall detection of subtle AD (whether related to a malignant or a benign structure). A systematic review using data from large screening trials reported that DBT detected an increase in cancers presenting as AD: 11.46 invasive cancers per 10,000 FFDM + DBT screening examinations compared to 3.63 invasive cancers detected per 10,000 FFDM screening examinations. This is likely due to the way DBT reduces masking of surrounding tissue and increases visibility of subtle but potentially important

breast changes like AD. Improved conspicuity with DBT may provide readers with better information to consider whether AD presentation is suspicious for breast cancer.

Evidence suggests that AD presentation visible on DBT but occult on FFDM or ultrasound should be treated as suspicious for cancer given the high PPV₃ and the number of missed cancers identified by AD presentation on DBT alone; however, AD can be a difficult breast cancer presentation to detect accurately.

DBT detects clinically relevant breast cancers presenting as AD (especially those that are mammographically occult) but it also represents benign final outcome findings and may contribute to the unnecessary work-up of benign outcomes

Research from symptomatic and asymptomatic populations highlights an issue with the unnecessary work-up of lesions with a final benign outcome and which may not have been clearly visible on FFDM imaging. Further research is needed to assess the sensitivity of DBT in distinguishing between malignant and benign structures if DBT is used in the assessment centre.

4.2.1. Detection of AD, which is often a subtle mammographic presentation for invasive breast cancer, increases with the use of DBT because of improved visibility

Systematic review and narrative literature reviews

Our search identified four narrative literature reviews (Michell & Batohi, 2018; Destounis et al., 2017; Mall et al., 2017; Peppard et al., 2015) and one pictorial atlas (Houssami et al., 2016) that presented high-level summaries about the role DBT can play in the detection of AD. The narrative reviews reported that, compared to FFDM, DBT:

- improves visibility of breast cancers presenting only as AD, reduces AD pseudo-effects and improves visibility and assessment of benign structures presenting as AD on FFDM
- supports better identification and classification of AD, particularly if the AD is FFDM or sonographically occult or is difficult to distinguish from focal asymmetry
- appears to detect AD in women with more dense breasts where the AD presentation is occult on FFDM or ultrasound, and
- improves readers' interpretive performance of ILC presenting as AD or spiculation.

All primary papers informing these literature reviews are discussed in this section.

Yun et al.'s systematic review (methodology described in *Chapter 4*) reported that FFDM + DBT resulted in the detection of more invasive breast cancers presenting as AD compared to FFDM: 11.46 invasive cancers per 10,000 FFDM + DBT screening examinations compared to 3.63 invasive cancers detected per 10,000 FFDM screening examinations. This data was drawn from the Malmö, STORM and OTS trials as well as two larger retrospective observational studies but no pooled analysis was completed due to small numbers of cancers and the heterogeneity issues in the imaging protocols and studies designs in the primary studies.

Prospective studies of women recalled to assessment

Data from one prospective study set in a screening program population suggested that DBT (either alone or with s2DM) detects more cases presenting as AD compared to FFDM. Caumo et al. (2018) reported on the imaging and histological characteristics of screen-detected cancers within a population-based screening program in Verona, Italy. Using images from the same dataset, 162 more AD lesions were seen on DBT + s2DM than seen on FFDM. The clinical significance of AD presentations was also investigated by Caumo et al. (2018). The authors found a statistically significant increase in the number of cancers presenting as AD detected with DBT

+ s2DM compared to those detected with FFDM: 35 cancers detected with DBT compared to five cancers detected by FFDM.

Retrospective studies of women recalled to assessment

Most of the evidence relating to the detection of AD presentation, breast cancer and women recalled to assessment is found in retrospective studies. The detail of these studies is provided in *Table 11*. The overall strength of evidence is limited by the small number of cancers presenting as AD but findings from the retrospective studies are consistent with those reported in the prospective studies.

Partyka et al. (2014) observed that because AD can be difficult to perceive, more cases of breast cancer presenting as AD could have been excluded from their study due to the inter-reader protocol used to determine included cases (i.e., consensus agreement for inclusion of cases). This limitation may apply to other studies as well and it also speaks to overall subtle lesion conspicuity with AD presentation.

In comparing mass characteristics in 548 women recalled to assessment with DBT, FFDM or other mammographic work-up views, Ni Mhuircheartaigh et al. (2017) found that two-thirds of breast cancers presenting as AD (n=13) were seen on both DBT and DSCV, four cancers were seen less clearly on DSCV and two cancers presenting as AD were only viewed on DBT. This finding resulted in a confident practice change in the population-based screening program to no longer complete spot compression views for AD presentation if DBT was performed as additional mammographic views provided no further relevant information compared to DBT view.

Increases in AD detection were also reported by McDonald et al. (2017), who found that in women recalled to assessment, AD was the main recall reason in 156 cases using DBT compared to 75 AD cases detected with FFDM ($p=.001$).

In Dibble et al.'s 2018 case-control study, reader confidence and inter-reader agreement for AD increased with the use of DBT compared to FFDM due to better visibility (discussed in *section 3.5.1*).

Prospective studies of mixed study populations (including symptomatic women only)

In Refaat & Matar's 2015 study, radiologists reported equivalent or superior visibility of AD when DBT was used (although this study was based on only six AD presentations). Skaane et al. (2012) also reported detection of one breast cancer presenting as AD on DBT that was occult on FFDM and ultrasound. While the overall numbers of cancers detected in each of these studies is small, the consistency in direction of effect suggests that DBT detects additional cancers presenting as subtle AD compared to FFDM.

Retrospective studies of mixed study populations (including symptomatic women only)

Five retrospective studies reported that DBT provided superior or at least equivalent visibility of AD presentations compared to FFDM, ultrasound or other mammographic views. These studies were usually based on readers' subjective assessment of lesion conspicuity of lesions worked up following a screening recall.

Mariscotti et al.'s 2017 analysis compared s2DM images to FFDM images, finding that visibility of AD was better with s2DM in approximately two-thirds of cases compared to FFDM and visibility was equivalent in the remaining one-third of AD cases. AD conspicuity was only better in FFDM for one AD presentation but no further information about this case was provided. Readers in Chan et al.'s 2017 study reported statistically significant results that breast cancers presenting as AD were better visualised on DBT compared to FFDM (increasing sensitivity by

21% with a very small decrease in specificity). Benign structures also had a higher mean score for conspicuity with DBT but no significance data was presented for these results. Yang et al. (2013) reported that almost all lesions presenting with AD were seen better on DBT compared to FFDM. Mariscotti et al. (2016) also reported that correct BIRADS lesion classification for AD mammographic presentation was significantly increased with DBT (65% with FFDM compared to 84% with DBT). In Mariscotti et al.'s 2017 study, AD was seen with equal or better visibility on s2DM in nearly 96% of cases ($p=.0009$).

Raghu et al. (2016) reported an increase in BIRADS 3 lesions (i.e., inconclusive screening mammographic results) following the implementation of DBT into their clinical practice and concluded that this was due to improved conspicuity and that diagnostic accuracy for AD was increased with DBT.

Reported PPV₃ rates vary but appear to be high when AD is only seen on DBT (i.e., mammographically occult):

- 53% for lesions detected only with DBT (Ray et al., 2015)
- 47% where AD is detected by DBT with no ultrasound correlate (Freer et al., 2015)
- 44% for biopsy when seen only on DBT and 21% for malignancy (Partyka et al., 2014).

4.2.2. DBT can detect AD that is occult on FFDM or ultrasound

Five studies investigated AD, reporting that DBT detects AD occult to FFDM or ultrasound.

Retrospective studies of women recalled to assessment

Partyka et al.'s 2014 retrospective observer study of DBT-suspicious but FFDM-occult AD reported on 26 breast cancers presenting as AD. The authors found that 73% of AD cases were only seen on DBT and 23% of cases were seen better on DBT compared to FFDM. Only one case was seen equally as well on both imaging modalities. Improved visibility of FFDM-occult AD was also demonstrated in Gilbert et al.'s 2015a retrospective analysis of TOMMY trial data, which found that AD/asymmetry was the dominant imaging feature in 42 cancers missed on FFDM compared to 37 cancers missed with either DBT or DBT + s2DM. This suggests that breast cancer presenting as AD or ASD is hard to detect, but their visibility improves with DBT.

Retrospective studies of mixed study populations (including symptomatic women only)

Like Partyka et al., Ray et al.'s 2015 review of pathologic features of breast cancers detected only with DBT. The authors reported that AD accounted for 74% of lesions seen only on DBT (i.e., FFDM-occult). Taskin et al. (2017) also reported that AD accounted for about 75% of lesions visible only with DBT.

Taskin et al. (2017), Partyka et al (2014) and Yang et al (2013) presented results by breast density where the AD presentation was FFDM-occult. Consistent results were reported: 79-84% of women with a breast cancer with AD presentation occult on FFDM or ultrasound had heterogeneously or extremely dense breasts (i.e., BIRADS 3 and 4). Ray et al (2015) reported that 63% of FFDM-occult AD lesions detected with DBT were found in women with more dense breasts. Most likely this speaks to both the difficulty of detection of AD overall and the impact of DBT.

4.2.3. DBT detects clinically relevant breast cancers presenting as AD (especially those that are mammographically occult) but it also represents benign final outcome findings and may contribute to the unnecessary work-up of benign outcomes

Not only do readers using DBT appear to be better able to detect subtle findings like AD, these findings may represent missed cancers if AD is seen only in the DBT view.

Retrospective studies of women recalled to assessment

In Partyka et al.'s study, nine of 19 lesions only seen as AD on DBT (i.e., FFDM-occult) were classified as BIRADS 4 or 5 (that is, suspicious for breast cancer). Four of these were breast cancers (two were IDC and two were DCIS) and two showed atypia. A further three invasive carcinomas (no further detail about type was provided) were seen on DBT and were then retrospectively confirmed on FFDM (i.e., the lesions could obviously be seen when the readers went back to the FFDM images). Partyka et al. reported a PPV₃ of 44%. The readers did not identify these areas as suspicious on FFDM and, without DBT, the cancers would not have been detected (i.e., these cancers represent missed cancers). The authors noted that the AD cases were drawn from prevalent screening and as such, may reflect higher levels of detection than would be seen in an incident screening round. Increased cancer detection presenting as FFDM-occult AD is also seen in other studies including Gilbert et al. (2015a), who reported a lower number of missed cancers with DBT (n=37) compared to FFDM (n=42) although information about prevalence/incidence was not available for this study.

Where AD was the dominant imaging feature, reported sensitivity increased in all retrospective studies when DBT is used (either instead of, with s2DM, or as an adjunct to) compared to FFDM. Reported sensitivity increases were all greater than 10 percentage points (Dibble et al., 2018; Mariscotti et al., 2017; Gilbert et al., 2015a). Dibble et al. (2018) also reported that, for confirmed AD cases, sensitivity in the detection of AD increased considerably for all readers:

- For attending readers with more than nine years breast imaging experience: an increase from 38% with FFDM to 97% with DBT images (95% CI: 0.38-0.39, 0.97-0.97; $p=0.001$), and
- For readers who were completing a one-year fellowship: an increase from 43% with FFDM to 75% with DBT (95% CI: 0.42-0.45, 0.67-0.81; $p=0.001$).

Mixed results were presented for specificity. One study reported increased specificity with DBT, one reported no significant difference between DBT and FFDM, and one study reported poorer specificity with s2DM compared to FFDM.

Given that AD may reflect a benign (rather than malignant) structure, increased detection of AD with DBT may result in more women presenting for further (and ultimately unnecessary) diagnostic work-up, leading to additional health sector costs and potential anxiety to women. Partyka et al. (2014) noted that the 22 cases of AD required additional diagnostic imaging and/or biopsy and surgical excision to determine that the AD presentation was associated with a benign breast structure rather than a malignancy. In 15 of these cases (i.e., the FFDM-occult presentations seen only on DBT), additional work-ups would have been avoided if DBT had not been used. This needs to be weighed against the four invasive cancers detected with DBT only that would have otherwise been missed.

Radial scar presenting as AD was noted by Destounis (2017), who reported that, in an internal review of radiographic findings from her clinical practice, the implementation of DBT resulted in additional excisions based on radial scar diagnosis (a finding that can be associated with breast cancer) but there was no malignancy at excision. She concluded that this may represent over-

diagnosis with the use of DBT + s2DM (which has algorithms that enhance AD detection). No specific study data for this finding was presented. Ni Mhuirheartaigh et al. (2017) discussed the impact of detecting benign radial scar presenting as a subtle AD, noting that it was more visible on DBT compared to DSCV: none of these cases were malignant but biopsy was performed on all to rule on breast cancer. Freer et al.'s (2015) study reported that 36 (100%) FFDM-occult cases for needle localisation and surgical excision were AD.

Ways to reduce the impact of diagnostic work-ups for benign structures presenting as AD were suggested by Taskin et al (2017) and Partyka et al. (2014): AD is more likely to represent a benign structure if it is seen on DBT but there is no correlate on ultrasound (which differs from Freer et al.'s reported PPV – see above); and visualise the AD presentation with ultrasound before progressing to biopsy.

Retrospective studies: unclear diagnostic pathway

In Bahrs et al. (2017) (which did not clearly describe the diagnostic pathway for study participants), there was a decrease in AD detected after the DBT reading, which was differs from the results presented in the other prospective studies. Four of the nine AD classified at FFDM as BIRADS 3 (inconclusive) retained the same classification with DBT but four were downgraded to BIRADS 1 or 2 and one was upgraded to BIRADS 4 with a benign finding. No further analysis or commentary is provided by the authors, so it is not possible to determine the clinical significance of these classification changes, but these findings are reflective of the issue of unnecessary work-up of AD presentations discussed in other studies.

Chan et al. (2017) discussed the increases in lesions classified as BIRADS 3 (i.e., inconclusive on mammogram imaging) or lower (normal or benign findings) seen with DBT. The authors reported that AD presentation detected with FFDM resulted in 22 BIRADS 3 classifications (inconclusive results) compared to only one AD detected with DBT receiving an inconclusive rating. Further, in this study more lesions suspicious for malignancy were detected with DBT. The authors concluded that conspicuity based on lesion shape and margin for AD associated with potential malignancy (or a benign structure) was improved with DBT. The authors also reported a mean conspicuity rating for AD of 5.8 with DBT compared to 3.3 with FFDM, ($p=.0002$)

Table 11: Primary studies investigating the role of DBT in detecting breast cancers presenting as subtle AD

11A Prospective studies of women recalled to assessment

Study	Sample	Study type	Findings
Caumo et al. (2018)	34,017 Italian women participating in a population-based screening program who had DBT compared to 29,360 women who had DM (historical control)	Prospective pilot evaluation embedded in a population-based screening program (Verona, Italy), using Hologic Dimensions unit	Total cancers detected: 468 (315 in the pilot, 153 from the control) <i>Lesion conspicuity where AD was the mammographic feature:</i> Detected with DBT + s2DM: 315 Detected with FFDM: 153 <i>Cancers presenting as AD detected with:</i> DBT + s2DM: 35 (11% of cancers detected with DBT+s2DM) FFDM: 5 (3.3% of all cancers detected with FFDM); $p=.005$

11B Retrospective studies of women recalled to assessment

Study	Sample	Study type	Findings
Dibble et al. (2018)	59 cases/controls imaged with bilateral FFDM and DBT screening images; cases drawn from reports containing words AD/possible AD	Retrospective case-control study (n=2) with an experimental using images obtained on a Hologic Selenia Dimensions unit	<p><i>Overall inter-reader agreement</i> FFDM images: Fair to moderate ($k=0.37$) DBT images: Moderate to good ($k=0.61$) <i>Reader confidence in detecting AD</i> FFDM images: 2.6 (on a scale of 0-4) DBT images: 3.2 (on a scale of 0-4); $p<.001$ <i>Sensitivity</i> With DBT: 0.59 (95% CI: 0.49, 0.69) With DM: 0.32 (95% CI: 0.26, 0.39), $p=.0006$ <i>Specificity</i> With DBT: 0.93 (95% CI: 0.93, 0.94) With DM: 0.93 (95% CI: 0.91, 0.95), $p=0.9997$</p>
McDonald et al. (2017)	1112 recalls from FFDM screening examinations 1366 recalls from DBT screening examinations	Retrospective reader (n=6) audit of population screening data using images from Hologic Dimensions unit	<p>Recalled finding of AD: FFDM images: 6.1% DBT images: 9.9%, difference 62.3% increase, $p<.001$ <i>Change in final BIRADS 3 finding/total findings</i> FFDM: 6.7% DBT: 4.5% <i>Difference</i>: 32.8% decrease, $p=.207$</p>
Ni Mhuircheartaigh et al. (2017)	548 Irish women recalled to assessment, 105 of whom were recalled because of AD; 74 of these women were imaged with DBT and DSVC	Retrospective reader (n=3) analysis of all women recalled to assessment over a six-month period	<p>Benign finding: 55/74 Malignant findings: - 12 invasive cancers - 2 DCIS - 5 radial scar with no evidence of malignancy 13/19 malignancies seen on spot compression, 4 malignancies presenting as AD were less clear, and 2 were not visible on spot compression</p>
Partyka et al. (2014)	All consensus-confirmed BIRADS 0 classifications where possible AD or AD was the reason for classification and when no other reason for AD existed (26/9982 FFDM and DBT screening examinations)	Single site, retrospective reader review (n=3) where readers were blind to diagnostic work-up/final pathology and bilateral screening images were obtained on a Hologic Selenia Dimensions unit	<p>51 cases of AD (total examinations = 9982); consensus agreement that AD present in 26 cases and not present in 25 cases. 19 AD cases (73%) were only seen on DBT. Of these, 7 were seen on subsequent spot compression views and 6 were seen on subsequent ultrasound views. Five cases also had MRI, which detected three of the AD cases. 9 AD lesions were classified as BIRADS 4 or 5 and 2 were classified as BIRADS 3. 6 AD cases (23%) cases were seen better on DBT. 1 AD case (4%) was seen equally as well on DBT and FFDM. Biopsy results from the 9 BIRADS 4 or 5 cases seen only on DBT: 2 invasive carcinomas; 2 DCIS; 3 radial scars; 2 lesions showing atypia. PPV₃ 21%; PPV₂ 44%. 16/19 women (84%) where AD was only seen on DBT had more dense breasts (BIRADS 3 or 4).</p>
Gilbert et al. (2015a) TOMMY trial	7060 women recalled to assessment following routine screening or because of family history, all of whom had bilateral DBT and FFDM imaging	Retrospective observer (n=26) study based on blinded review of a cancer-rich cohort with DBT images from Hologic Selenia Dimensions unit	<p>110 cancers presenting as AD or ASD/1863 AD or ASD presentations <i>Sensitivity where dominant imaging feature was AD or ASD</i>: FFDM + DBT: 82% DBT + s2DM: 82% FFDM: 71% <i>Specificity where dominant imaging feature was AD or ASD</i>: FFDM + DBT: 75% DBT + s2DM: 76%, $p=.001$ FFDM: 64% Missed cancers with DBT or s2DM: 37 Missed cancers with FFDM: 42</p>

Study	Sample	Study type	Findings
Lourenco et al. (2014)	FFDM: 1175 FFDM + DBT: 827	Retrospective review of two cohorts (DBT alone=2012/13, FFDM=2011/12), single reading with CAD. FFDM performed using GE Senographe series. DBT performed with Hologic Selenia Dimensions system.	Recalled to assessment: DM: 7 (0.6% of primary mammographic findings) DBT: 44 (5.3% of primary mammographic findings); $p<.0001$

11C Prospective studies of mixed study populations (asymptomatic and symptomatic women)

Study	Sample	Study type	Findings
Skaane et al. (2012)	50 Norwegian women attending a breast imaging clinic for complete conventional imaging work-up and needle biopsy (if indicated) + DBT	Prospective, subjective side by side feature analysis using Selenia Dimensions unit for DBT and GE Senographe unit for FFDM images	Total cancers detected: 23 Cancer detected on DBT alone presenting as AD: 1

11D Retrospective studies of mixed study populations (asymptomatic and symptomatic women)

Study	Sample	Study type	Findings
Taskin et al. (2017)	107 cases with suspicious findings on DBT (i.e., AD presentation not seen on ultrasound or FFDM)	Retrospective evaluation of records, biopsy, or FFDM/US results; DT/FFDM imaging with Siemens Mammomat Inspirations unit	Women with BIRADS 3: 52% Women with BIRADS 4: 32% 74% of suspicious findings were AD (79/107) 50/79 (63%) of AD presentations had an MRI finding with 14 being malignancies and 36 benign lesions. All malignancies were invasive: 11 IDC and 3 ILC.
Ray et al. (2015)	268 consecutive BIRADS category 4 or -5 lesions imaged with both FFDM and DBT (combo mode)	Retrospective single site reader review using prevalent bilateral screening images obtained on a Hologic Selenia Dimensions unit	19 lesions were FFDM-occult. 14/19 lesions were AD seen only on FFDM + DBT. 12/19 FFDM-occult lesions (63%) were in dense breasts (BIRADS 3 or 4) PPV ₃ : 53% for all mammographically occult lesions

11E Prospective studies of symptomatic women only

Study	Sample	Study type	Findings
Refaat & Matar (2015)	25 Egyptian women with breast cancer imaged by DBT and FFDM	Prospective case study using GE Senographe Essential unit	Total malignancies presenting as AD: 5 (one AD was FFDM-occult) <i>Equivalence rating based on reader's subjective view:</i> DBT superior view compared to FFDM: 2 cases DBT equivalent view compared to FFDM: 3 cases DBT inferior view compared to FFDM: 0 cases

11F Retrospective studies of symptomatic women only

Study	Sample	Study type	Findings
Mariscotti et al. (2016)	Participants: 83 consecutive women with newly diagnosed ILC (107 ILC proven at histology) Unit: Hologic Selenia Dimensions	Retrospective, multi-reader (n=12) study	Correct classification of mammographic findings FFDM: 65% FFDM + DBT: 84%, $p < .0001$ <i>Number of cases: Reader A</i>
Raghu et al. (2016)	All diagnostic mammograms in the 12 months before implementation of DBT and for three consecutive years post-implementation: - 3576 FFDM images (Y0) - 3394 DBT images (Y1) - 4541 DBT images (Y2) - 4676 DBT images (Y3)	Retrospective study to evaluate effect of DBT in diagnosis on final BIRADS assessment categories using Hologic Dimensions unit	Y0: 17/1374 (1.2%) lesions classified as BIRADS 3 Y3: 21/901 (2.3%) lesions classified as BIRADS 3

11G Prospective studies: unclear diagnostic pathway

Study	Sample	Study type	Findings
Bahrs et al. (2017)	105 consecutive patients with non-calcified BIRADS 3 lesions with no ultrasound correlate	Prospective reader (n=3) study in which participants had DBT not DSCV or magnification Selenia Dimensions	<i>Number of cases: Reader A</i> FFDM: 12; DBT: 8 <i>Number of cases: Reader B</i> FFDM: 4; DBT: 6 <i>Number of cases: Reader C</i> FFDM: 13; DBT: 7

11H Retrospective studies: unclear diagnostic pathway

Study	Sample	Study type	Findings
Chan et al. (2017)	Participants: 134 women imaged with wide-angle DBT (60°) compared to diagnostic work-up FFDM images; non-	Retrospective reader (n=6) study of enriched dataset to compare FFDM to DBT using a GE prototype (GEN2)	61 malignant lesions and 74 benign cases with a total of 14 AD presentations. <i>AD lesions classified as BIRADS 3 and below</i> FFDM images: 22 malignant, 1 benign DBT images: 1 malignant, 0 benign <i>AD lesions classified as BIRADS 4a and above</i>

Study	Sample	Study type	Findings
	calcified lesions only	DBT)	FFDM images: 56 malignant, 5 benign DBT images: 77 malignant, 6 benign <i>Mean conspicuity for malignancies presenting as AD (scale 1-10)</i> FFDM images: 3.3 DBT images: 5.8 ($p=.0002$) <i>Mean conspicuity for benign structures presenting as AD (scale 1-10)</i> FFDM images: 2.5 DBT images: 4.0 (p -value not provided)
Mariscotti et al. (2017)	231 Italian women with 250 biopsy-proven suspicious breast lesions (consecutive dataset) imaged with FFDM + DBT (combo mode)	Retrospective analysis in which radiologists (n=2) blinded to histology reviewed s2DM and FFDM using Hologic Selenia Dimensions unit with s2DM images reconstructed using C-view software	32/250 (12.8%) of lesions presented with AD. <i>Lesion conspicuity where AD was the mammographic feature:</i> AD seen better with s2DM reconstructed image: 64.5% AD seen better with FFDM: 3.2% Equal visibility: 32.3%, $p=.0009$ <i>Concordance between s2DM and FFDM for AD (BIRADS assessment)</i> Cohen's k (95% CI): 0.36 (95% CI: 0.16, 0.56) <i>AUC</i> s2DM: 0.69 ± 0.10 FFDM: 0.67 ± 0.10 , $p=.89$ 54.5% of participants had less dense breasts; 45.5% of patients had more dense breasts (BIRADS 3 or 4).
Takamoto et al. (2013)	195 breast images (from 99 Japanese women) diagnosed with breast cancer imaged with FFDM and DBT	Retrospective reader (n=2) analysis using Hologic Selenia Dimensions unit	18 breast cancers presenting as AD detected Following DBT in affected breast, increase in BIRADS category for eight (no further data provided): OR=6.91, (95% CI: 1.7-, .57; $p=.006$) and a decrease in category for 3/5 of unaffected breasts ($p=.001$): not detected with DBT but seen on FFDM.
Yang et al. (2013)	59 pathologically proven breast cancers imaged with FFDM + DBT (combo mode)	Retrospective non-blinded reader (n=3) simultaneous viewing of DBT and DM images using Hologic's Selenia Dimensions	6 lesions presented as AD (10.2%) <i>Lesion conspicuity where AD was the mammographic feature:</i> AD seen somewhat better on FFDM + DBT: 27.8% AD seen better on FFDM + DBT: 66.7% Same: 5.5% 79% of participants had more dense breasts (BIRADS 3 or 4)

4.3. Focal asymmetry

Asymmetries can be focal (which can be indicative of malignancy) or global (which is more likely to reflect a normal breast structure unless the degree of difference in asymmetry is very large). A focal asymmetry is a small area of fibroglandular tissue that is visible on one projection and in one breast. It is described in the BIRADS lexicon as being a planar, interspersedly fatty and lacking 3D conspicuity and convexity. Eleven studies and three narrative literature reviews included in this literature review commented on asymmetry. A summary of findings is included in *Table 12 (A-E)*.

Systematic reviews and narrative literature reviews

Systematic reviews: none

Narrative literature reviews: Eghtedari et al. (2018); Destounis (2017); Mall et al. (2017)

RCTs and prospective studies

RCTs: none

Three prospective studies:

Women recalled to assessment and symptomatic women: Seo et al. (2016)

Only symptomatic women and/or women with a diagnosed breast cancer: Refaat & Matar (2015)

Diagnostic pathway unclear: Bahrs et al. (2017)

Retrospective studies

Eight retrospective studies:

Women recalled to assessment: McDonald et al. (2017); Lourenco et al. (2014)

Only symptomatic women and/or women with breast cancer: Mariscotti et al. (2016); Raghu et al. (2016); Takamoto et al. (2013); Yang et al. (2013)

Diagnostic pathway unclear: Chan et al. (2017); Mariscotti et al. (2017)

Key findings

Asymmetries are a subtle mammographic finding which can be focal (likely to represent a breast cancer) or global (more likely to represent a benign breast structure). Increased lesion conspicuity and reduced tissue overlap with DBT compared to FFDM means that readers may be more confident in dismissing as benign, asymmetries which may look suspicious on DM.

All studies (almost all of which were in mixed study populations) reported improved conspicuity of asymmetry with DBT: two studies reported superior visibility in 75% of cases but did not describe any further implications. Other studies demonstrated that DBT or s2DM had superior sensitivity for asymmetry compared to FFDM (with incremental increases ranging from 6% to 27.5%). Asymmetry as a primary mammographic finding decreased with the use of DBT, with one larger study based on lesions recalled to assessment indicating a decrease in recall with DBT of approximately 50%. Similarly, reported downgrading in BIRADS 3 classifications (i.e., inconclusive findings were resolved to normal or benign findings) was approximately 45% and were a result of improved confidence to dismiss an area as benign/normal or reclassification of the finding as a mass due to better visibility of the lesion margin and shape.

4.3.1. Use of DBT results in reclassification of primary mammographic findings to suspicious masses OR more benign/normal, which may contribute to the unnecessary work-up of benign outcomes

Narrative literature reviews

Our search identified three narrative literature reviews (Eghtedari et al., 2018; Destounis, 2017; Mall et al., 2017), which presented high-level summaries about the role DBT can play in the detection of focal asymmetries. The narrative reviews reported that compared to FFDM, DBT:

- reduces tissue overlap making it easier to determine if breast structures are benign, and
- results in differential classification of lesions to either benign (with no need for additional work-up views) or suspicious masses.

Limited further description was provided in the narrative literature reviews, and all studies summarised in these literature reviews are discussed below.

Prospective studies of mixed study populations (including symptomatic women only)

Three prospective studies reported on the impact of DBT on asymmetry as a mammographic finding. Bahrs et al. (2017) reported a statistically significant decrease of 44.8% in lesions classified as asymmetry when DBT images were read compared to FFDM ($p<.0001$). Given that the study population was women with lesions classified as BIRADS 3 (inconclusive) at FFDM imaging, 39/57 asymmetries were down-graded to BIRADS 1 or 2 (benign or normal) and only one was upgraded to BIRADS 4 which on histology was a carcinoma no special type following assessment with DBT. No downgraded lesions had malignant findings at follow-up (term not described). While a small study, the authors concluded that DBT resulted in reduced follow-up of indeterminate lesions presenting as BIRADS 3 asymmetry. In a very small study by Refaat & Matar (2015), DBT provided superior visibility of asymmetry in 75% of asymmetry presentations.

One other recent prospective study (Seo et al., 2016) reported on features of cancers occult to both FFDM and DBT. A total 129 cancers were detected (113 invasive cancers and 16 DCIS). Of these 48 cancers (37 invasive cancers and 11 DCIS) were unable to be detected on FFDM but could be seen when FFDM + DBT. No further detail about histology was provided for these cancers; however, of the cancers missed on both imaging modalities, seven were in women with extremely dense breast tissue, eight were asymptomatic and eight were IDC and two were DCIS. The authors reported that cancers were still difficult to detect due to a mammographic presentation based on focal asymmetry or negative mammographic findings.

Retrospective studies of women recalled to assessment

One study reported changes with the use of DBT that resulted in different case management. Using data from a screening population, Lourenco et al. (2014) reported that after the implementation of DBT, 827 lesions were recalled to assessment. Prior to implementation (1175 lesions), 32.2% of recalls to assessment were for asymmetry and a further 32.3% were recalled for focal asymmetry. Implementation of DBT resulted in a substantial reduction in recalls for asymmetry as a primary mammographic finding: 13.3% for asymmetry and 18.3% for focal asymmetry ($p<.0001$). Reasons for this were improved lesion margin visibility resulting in either downgrading of areas suspicious on FFDM to benign/normal findings or reclassification to masses. It is not clear whether these study result included a mix of prevalent and incident screening examinations but given that data is drawn from a screening population, it is likely to be more representative of real-world clinical practice than some other studies reported in this literature review and it is indicative of the potential of reduced future work-up due to increased reader confidence.

Retrospective studies of mixed study populations (including symptomatic women only)

The seven retrospective studies (mixed study populations or symptomatic women only) reported consistent information about how asymmetries are visualised with DBT compared to DM.

Three studies reported on overall conspicuity, which increased in all studies when DBT was used. For example, in Chan et al.'s (2017) fully crossed reader study, conspicuity (rated by six radiologists with median experience of 20 years) was much higher for malignant asymmetries when seen on DBT compared to DM (5.2 compared to 3.1; $p=.00001$). Mariscotti et al. (2017) reported that s2DM had superior visibility compared to FFDM in 42.1% of cases and equal visibility in 31.6% of cases ($p=.0009$). Yang et al. (2013) reported better visibility of asymmetry in more than 80% of asymmetries.

Improved conspicuity also led to changes in BIRADS classification (i.e., a reduction in the number of BIRADS ≥ 3 classifications with a primary mammographic finding of asymmetry or a change in primary mammographic finding to a mass). In Chan et al.'s (2017) study, sensitivity improved 27.5% for asymmetries as did specificity when DBT was used. Mariscotti et al. (2016) reported increased accuracy of classification compared to proven histology, with an 11% improvement with DBT. Mariscotti et al. (2017) also reported increased sensitivity and specificity with s2DM compared to FFDM (a non-significant increase of 6% for sensitivity and a non-significant increase of 8% in specificity). These results suggest that DBT or s2DM will identify more cancers presenting as asymmetries and correctly dismiss more women without breast cancer compared to FFDM.

Other smaller studies also demonstrated decreases in recall to assessment due to asymmetry, but the findings were considerably smaller (for example, McDonald et al. reported a decrease of approximately 20% and Raghu et al. reported a decrease of 30%).

Table 12: Primary studies investigating the role of DBT in detecting breast cancers presenting as asymmetry

12A Retrospective studies of women recalled to assessment

NB: Study design described previously in Table 11

Study	Findings	
Lourenco et al. (2014)	Recalled to assessment with DM: 32.2 and 32.3 (64.5% in total)	Recalled to assessment with DBT: 13.3 and 18.3, (31.6%) $p < .001$
McDonald et al. (2017)	<i>Change in recalled primary mammographic finding</i> FFDM: 53.0% DBT: 44.9% Difference: 15.2 percent decrease, $p < .001$	<i>Change in final BIRADS 3 finding/total findings</i> FFDM: 10.9% DBT: 13.1% Difference: 20.2% increase, $p = .207$

12B Prospective studies of mixed study populations (including asymptomatic and symptomatic women)

NB: Study design described previously in Table 11

Study	Findings		
Bahrs et al. (2017)	<i>Number of cases: Reader A</i> FFDM: 56 DBT: 16	<i>Number of cases: Reader B</i> FFDM: 63 DBT: 27	<i>Number of cases: Reader C</i> FFDM: 51 DBT: 12

12C Retrospective studies of mixed study populations (including asymptomatic and symptomatic women)

NB: Study design described previously in Table 11

Study	Findings	
Chan et al. (2017)	<i>BIRADS ≤ 3 malignancies</i> FFDM: 36 DBT: 8 <i>BIRADS ≤ 3 benign lesions</i> FFDM: 57 DBT: 59	<i>BIRADS $\geq 4a$ malignancies</i> FFDM: 66 DBT: 94 <i>BIRADS $\geq 4a$ benign lesions</i> FFDM: 33 DBT: 31
Mariscotti et al. (2017)	27/250 (10.8%) of lesions presented with asymmetry <i>Lesion conspicuity where asymmetry was the mammographic feature</i> Seen better with s2DM reconstructed image: 42.1%	<i>Sensitivity for asymmetry</i> s2DM: 93% FFDM: 87%, $p > .99$ <i>Specificity for AD</i> S2DM: 83% FFDM: 75% $p > .99$

Study	Findings
	<p>Seen better with FFDM: 26.3%</p> <p>Equal visibility: 31.6%, $p=.0009$</p> <p>Concordance between s2DM and FFDM (BIRADS assessment)</p> <p>Cohen's k (95% CI): 0.83 (95% CI: 0.68, 0.99)</p> <p>AUC</p> <p>s2DM: 0.93 ± 0.05</p> <p>FFDM: 0.83 ± 0.08, $p=.29$</p>

12D Prospective studies of symptomatic women only

NB: Study design described previously in Table 11

Study	Findings
Refaat & Matar (2015)	<p>Total malignancies presenting as focal density: 4; <i>Equivalence rating based on reader's subjective view:</i></p> <p>DBT superior view compared to FFDM: 3 cases; DBT equivalent view compared to FFDM: 1 case; DBT inferior view compared to FFDM: 0 cases</p>

12E Retrospective studies of symptomatic women only

NB: Study design described previously in Table 11

Study	Findings
Mariscotti et al. (2016)	<p>Correct classification of mammographic findings</p> <p>FFDM: 66%</p> <p>FFDM + DBT: 77%, $p<.0001$</p>
Raghu et al. (2016)	<p>Y0: 512/1374 (37%) lesions classified as BIRADS 3</p> <p>Y3: 245/901 (27%) lesions classified as BIRADS 3</p>
Takamoto et al. (2013)	<p>10 breast cancers. Following DBT in affected breast, increase in BIRADS category for eight (no further data provided): OR=9.32, (95% CI: 1.91, 45.36; $p=.006$) and a decrease in category for 3/5 of unaffected breasts ($p=.001$): not detected with DBT but seen on FFDM.</p>
Yang et al. (2013)	<p>12 lesions presented as density (20.3); <i>Lesion conspicuity where density was the mammographic feature:</i></p> <p>Seen somewhat better on FFDM + DBT: 41.7%; Definitely seen better on FFDM + DBT: 41.7%; Same: 16.6%</p>

4.4. Microcalcifications

Specific patterns of microcalcifications (such as linear or branching microcalcifications) are a key mammographic presentation for breast cancer, particularly DCIS. Calcifications have high x-ray attenuation properties meaning that some patterns of microcalcifications suspicious for breast cancer are usually easy to detect on DM. Mixed results have been reported on DBT's ability to depict microcalcifications as clearly as FFDM. This may be because, unlike masses, AD or asymmetries, microcalcifications tend to be less affected by tissue overlap and are therefore adequately visualised and easily detected on FFDM. Complicating matters is the fact that presentation of other micro- or macrocalcification patterns (such as oval) can be indicative of benign or normal breast structures. This literature review describes findings from 17 primary studies and two narrative literature reviews that reported on detection of microcalcifications with DBT (either as a primary study outcome or within a wider radiologic feature analysis). The studies discussed in this literature review are listed below. A summary of findings is included in Table 13 (A-F).

Systematic reviews and narrative literature reviews

Systematic reviews: none

Two narrative literature reviews: Johnson (2017); Mall et al. (2017)

RCTs and prospective studies

RCTs: none

Two prospective studies:

Women recalled to assessment: Caumo et al. (2018)

Only symptomatic women and/or women with a diagnosed breast cancer: Refaat & Matar (2015)

Retrospective studies

15 retrospective studies:

Women recalled to assessment: McDonald et al. (2016); Gilbert et al. (2015a); Lourenco et al. (2014); Morel et al. (2014); Kopans et al. (2011)

Women recalled to assessment and symptomatic women: Spangler et al (2011)

Only symptomatic women and/or women with a diagnosed breast cancer: Mariscotti et al. (2016); Raghu et al. (2016); Cai et al. (2015); Takamoto et al. (2013); Yang et al. (2013)

Diagnostic pathway unclear: Mariscotti et al. (2017); Berger et al. (2016); Clauser et al. (2016); Tagliafico et al. (2015b); Destounis et al. (2013)

Key findings

Mixed results are presented for microcalcifications: early studies suggested that DBT was an inferior imaging technique for microcalcifications, but later studies report equivalence

Specific patterns of microcalcifications are a key mammographic presentation for breast cancer, particularly DCIS. Calcifications in general have high x-ray attenuation properties meaning that microcalcifications suspicious for breast cancer are usually easy to detect on DM. Microcalcifications with patterning suspicious for breast cancer may be more difficult to detect on DBT due to the way that images are viewed (1mm slices compared to a larger 'slabbing' technique), or due to engineering issues in different DBT-capable units such as resolution.

Bearing in mind that DCIS is easily seen on FFDM, overall, studies report mixed results on whether DBT (either alone or as FFDM + DBT) has equivalent or inferior performance in terms of detecting microcalcifications compared to FFDM. There is some consensus that DBT alone may not be sufficient for the detection of cancers presenting with microcalcification as the primary mammographic finding. There is consensus that image quality is now equivalent to or in some cases better than FFDM for microcalcifications. s2DM + view DBT images as a slab shows some promise in improving DBT's performance but further evidence of success is required.

4.4.1. There is mixed evidence about DBT's equivalence in detecting suspicious microcalcifications compared to DM which may reflect the ease at which suspicious micro calcifications are detected on FFDM and/or algorithm improvements

Narrative literature reviews

Johnson (2017) briefly described the mixed evidence about the role of DBT in detecting breast cancers presenting as microcalcifications. The authors linked differences in study findings to:

- i) the fact that lower sensitivity and specificity for the identification of breast cancers presenting as microcalcifications may reflect the already high detection of these cancers due to their high contrast on x-ray
- ii) the engineering and technical differences in different DBT units which may affect the visibility of microcalcifications on DBT images, and
- iii) viewing 1mm slices, which may reduce readers' ability to detect suspicious patterns/clusters of microcalcifications compared to FFDM images as obvious clustering may be reduced (this issue could be overcome by slabbing techniques).

All the studies reported in this section were completed using Hologic's Selenia Dimensions units (except for Kopans et al., 2011), so any issues related to equipment are likely to be minimised.

Our search identified one other narrative literature review that presented a high-level summary on the role DBT can play in the detection of microcalcifications (Mall et al., 2017). Mall et al. reported that DBT can result in some loss of image characteristics (depending on the settings) but that there are ways to reduce noise to improve visibility and that there are ways to reduce some of the issues with microcalcification visibility. The authors also noted that there have been mixed results overall in terms of sensitivity, AUC measurement and lesion conspicuity for microcalcification presentation.

Prospective studies of women recalled to assessment

Caumo et al. (2018) completed a comparative analysis (n=314 cancers) using prospective data from the Verona screening program to describe the characteristics of cancers detected with DBT + s2DM compared to retrospective control data with FFDM screening. No significant increase in cancers presenting as microcalcifications was reported with DBT + s2DM compared to FFDM (13.7% of total cancers compared to 19.1%, not significant) and a borderline significant decrease in masses presenting with microcalcifications (8.9% of total cancers compared to 14.5%). The authors provided no further comments about the significance of this finding.

Prospective studies of symptomatic women only

Refaat & Matar reported that image quality for microcalcifications was better or equivalent with DBT compared to FFDM, a finding reflected by Kopans et al. (2011), Destounis et al. (2013) and Mariscotti et al. (2016) (see below).

Retrospective studies of symptomatic women only

In studies of symptomatic women only, Mariscotti et al. (2016) reported that FFDM + DBT resulted in more accurate BIRADS classification of microcalcifications (72% of findings were correct compared to 65% with FFDM, $p < .0001$). This may be due to the use of the combo mode (rather than comparing DBT alone to FFDM). Yang et al. (2013) reported that the level of improvement in FFDM + DBT imaging for microcalcifications was much lower compared to the improvement seen for other features (7/11). FFDM + DBT classified 23/27 asymmetric

densities, 12/12 AD and 12/12 masses as BIRADS 4 classification or higher when these were scored as BIRADS 0 with FFDM.

Retrospective studies of mixed study populations (including symptomatic women only)

Some of the mixed results reported for the sensitivity and specificity of microcalcification detection with DBT may be due to the use of early DBT-capable units (either prototypes, technical issues such as longer acquisition time with earlier units which could have increased motion noise, or earlier versions of reconstruction algorithms: these could have made it more difficult to detect microcalcifications). One of the earliest studies retrieved for this literature review was Spangler et al (2011). In their retrospective review of 100 paired FFDM and DBT examinations, the authors reported that sensitivity for microcalcification detection was higher with FFDM compared to DBT (85% compared to 75%); however, ROC analysis reported only a slightly higher AUC measurement for FFDM (0.76, SD = 0.03 compared to 0.72, SD = 0.04). BIRADS agreement favoured FFDM with nine breast cancers assigned a BIRADS < 2 classification compared to three cases viewed with DBT. FFDM was also more accurate at classifying benign lesions. It is important to note that Spangler et al.'s study was completed on an early Hologic unit prior to FDA approval (in 2008/09). The authors concluded that FFDM was slightly more sensitive for cancers presenting as microcalcifications compared to DBT alone overall.

Rafferty et al. (2014) reported AUC measurement for lesions presenting with microcalcifications or non-calcified lesions (i.e., masses, AD and asymmetries). They reported that AUC increased for all lesion types with or without microcalcifications but the greatest gain favoured FFDM + DBT for non-calcified lesions (a statistically significant gain of 0.088, which was twice that of the reported gain reported for FFDM + DBT lesions with microcalcifications). Cai et al. noted no specific issues in visualising microcalcifications with DBT but provided no further comment.

Retrospective studies: diagnostic pathway unknown

In Tagliafico et al.'s more recent study (2015b), the authors raised concerns that DBT may have missed four malignant and high-risk lesions that had been detected via microcalcification presentation with FFDM.

Since the publication of the studies previously discussed, a range of other studies reporting equivalent or superior diagnostic performance for microcalcifications have been completed. Different dimensions of diagnostic performance that have been explored include the following:

- Clauser et al. (2016) reported AUC measurement, sensitivity and diagnostic accuracy for microcalcification presentation were all similar for both modalities (bearing in mind that DBT was used in wide-angle) and concluded that results may support DBT as a stand-alone technique.
- In contrast to Clauser et al. (2016), Morel et al. (2014) provided commentary that there were no significant differences in detection or the sensitivity of DBT for lesions presenting as microcalcifications compared to FFDM but that DBT alone may provide insufficient imaging if used alone.

In Mariscotti et al.'s 2017 study, DBT + s2DM and FFDM had much lower (but comparable) AUC measurement for microcalcifications and AD compared to masses and asymmetries. Gilbert et al. (2015a) presented TOMMY trial AUC measurement by whether microcalcifications are present. There was no statistical difference in diagnostic accuracy for FFDM + DBT and DBT + s2DM compared to FFDM alone: FFDM + DBT and s2DM and DBT + s2DM had slightly better diagnostic accuracy (0.75 or 0.74) compared to FFDM (0.73) although statistical significance was not achieved.

4.4.2. Image quality is not inferior with DBT

Microcalcifications are clearly visible on FFDM. Two important early studies assessed whether image quality of microcalcifications was superior with DBT. Kopans et al. (2011) used a prototype GE DBT-capable unit to compare DBT_{MLO} to FFDM. In this study, the authors retrospectively reviewed 119 cases with suspicious microcalcification presentations to determine how easy it was to detect microcalcifications using readers' subjective assessment on a scale of 1 to 5 (with 1 being seen much more clearly on FFDM and 5 being seen much more clearly on DBT). Both readers were experienced but there were considerable differences in the depth of experience (having five- or 35-years' experience). Overall, readers considered that microcalcifications were seen more clearly on 41.6% of the DBT images compared to FFDM images, were about the same in 50.4% of cases and superior on FFDM in 8% of cases. Reader agreement was excellent (kappa statistic = 0.9559). Destounis et al. (2013) completed a single institution comparison of the visualisation and image quality of 177 examinations presenting with microcalcifications in FFDM and DBT, reporting that (like Kopans et al.), all microcalcifications were seen with DBT and that the image quality was equivalent or superior in 92.2% of cases, with similar performance for both malignant and benign presentations. Finally, very similar results were presented by Mariscotti et al. (2017), who reported that microcalcifications were seen with equal or better visibility on s2DM compared to FFDM in over 94% of cases ($p=.0009$), with a high degree of concordance with BIRADS assessment and increased sensitivity. These later results contrast with Spangler et al.'s earlier findings of lower visibility with DBT compared to FFDM.

Table 13: Primary studies investigating the role of DBT in detecting breast cancers presenting as microcalcifications

13A Prospective studies of women recalled to assessment

NB: Study design described previously in Table 11

Study	Study details	Findings
Caumo et al. (2018)	Prospective pilot evaluation embedded in a population-based screening program (Verona, Italy). 34,017 Italian women participating in a population-based screening program who had DBT compared to 29,360 women who had DM (historical control) Unit: Hologic Selenia Dimensions	<i>Cancers presenting as microcalcifications detected with:</i> DBT + s2DM: 43(13.7% of cancers detected with DBT+s2DM) FFDM: 29 (19.1% of all cancers detected with FFDM); $p=.13$

13B Retrospective studies of women recalled to assessment

NB: Study design described previously in Table 11

Study	Study details	Findings
McDonald et al. (2017)	Retrospective reader (n=6) audit of women with BIRADS 3 at diagnostic mammogram Participants: FFDM: 1112 women and DBT: 1366 women with similar baseline characteristics Unit: Hologic Selenia Dimensions	<i>Change in recalled primary mammographic finding</i> FFDM: 15.9% DBT: 15.9% <i>Difference: No change</i> <i>Change in final BIRADS 3 finding/total findings</i> FFDM: 34.2% DBT: 26.9% <i>Difference: 21.3% decrease, $p=.097$</i>

Study	Study details	Findings
Gilbert et al. (2015a)	Retrospective, multicentre (n=6), blinded multi-reader (n=26) matched study (TOMMY trial); cancer-enriched cohort; prospective data collection Participants: 7060 women recalled to assessment or aged <50y with a family history of breast cancer (cohort=8869); exclusions not described; 1137 malignancies (788 IDC; 109 ILC; 203 DIC; 59 other); 5691 normal Unit: Hologic Selenia Dimensions and C-view software <i>TOMMY trial: also cited in Phi et al. (2018)</i>	283 cancers presenting as microcalcifications <i>Sensitivity</i> FFDM + DBT: 88% DBT + s2DM: 85% FFDM: 88% <i>Specificity</i> FFDM + DBT: 39% DBT + s2DM: 44%, $p=.001$ FFDM: 31%
Lourenco et al. (2014)	Retrospective review of two cohorts (DBT alone=2012/13, FFDM=2011/12), single reading with CAD. FFDM performed using GE Senographe series Participants: FFDM: 1175; FFDM + DBT: 827 Unit: Hologic Selenia Dimensions	Recalled to assessment with DM: 158 (13.5%) Recalled to assessment with DBT: 168 (20.3%)
Kopans et al. (2011)	Retrospective study with blinded analysis Participants: 119 cases with FFDM and DBT images showing microcalcifications Unit: GE prototype	Reader 1: 42% more visible with DBT; 50.4% seen equally clearly on FFDM and DBT; 7.6% seen more clearly on FFDM Reader 2: 41.1% more visible with DBT; 50.4% seen equally clearly on FFDM and DBT; 8.4% seen more clearly on FFDM; kappa statistic = 0.9559

13C Retrospective studies of mixed study populations (including asymptomatic and symptomatic women)

NB: Study design described previously in Table 11

Study	Study details	Findings
Spangler et al. (2011)	Retrospective, multimodal paired analysis by 5 readers Participants: 100 paired FFDM and DBT examinations (20 malignancies, 40 benign; 40 screening cases randomly selected) Unit: research unit (not further described)	<i>Sensitivity overall</i> FFDM: 84% (95% CI: 79, 88) DBT: 75% (95% CI: 70, 80) <i>Sensitivity malignancies and benign lesions</i> FFDM: 90% / 81% DBT: 80% / 68% <i>AUC</i> FFDM: 0.76 \pm 0.03 DBT: 0.72 \pm 0.04

13D Prospective studies of symptomatic women only

NB: Study design described previously in Table 11

Study	Study details	Findings
Refaat & Matar (2015)	Prospective case study Participants: 25 Egyptian women with breast cancer imaged by DBT and FFDM Unit: GE Senographe Essential unit	Total malignancies presenting as microcalcifications: 11 <i>Equivalence rating based on reader's subjective view:</i> DBT superior view compared to FFDM: 8 cases DBT equivalent view compared to FFDM: 3 cases DBT inferior view compared to FFDM: 0 cases

13E Retrospective studies of symptomatic women only

NB: Study design described previously in Table 11

Study	Study details	Findings
-------	---------------	----------

Study	Study details	Findings
Mariscotti et al. (2016)	Retrospective, multi-reader (n=12) study Participants: 83 consecutive women with newly diagnosed ILC (107 ILC proven at histology) Unit: Hologic Selenia Dimensions	Correct classification of mammographic findings FFDM: 65% FFDM + DBT: 72%, $p < .0001$
Raghu et al. (2016)	Retrospective study to evaluate effect of DBT in diagnosis on final BIRADS assessment categories Participants: all diagnostic mammograms in the 12 months before implementation of DBT and for three consecutive years post-implementation: - 3576 FFDM images (Y0) - 3394 DBT images (Y1) - 4541 DBT images (Y2) - 4676 DBT images (Y3) Unit: Hologic Dimensions unit	Y0: 545/1374 (39.7%) lesions classified as BIRADS 3 Y3: 419/901 (46%) lesions classified as BIRADS 3
Takamoto et al. (2013)	Retrospective reader (n=2) analysis Participants 195 breast images (from 99 Japanese women) diagnosed with breast cancer imaged with FFDM and DBT Unit: Hologic Selenia Dimensions	44 breast cancers Following DBT in affected breast, increase in BIRADS category for eight (no further data provided): OR=0.61, (95% CI: 0.18, 2.02; $p = .415$).
Yang et al. (2013)	Retrospective non-blinded reader (n=3) simultaneous viewing of DBT and DM images Participants: 59 pathologically proven breast cancers imaged with FFDM + DBT (combo mode) Unit: Hologic's Selenia Dimensions	23 lesions presented as microcalcifications (20.3%) <i>Lesion conspicuity where microcalcification was the mammographic feature:</i> Seen somewhat better on FFDM + DBT: 5.8% Definitely seen better on FFDM + DBT: 5.8% Same: 88.4%

13F Retrospective studies: diagnostic pathway unknown

NB: Study design described previously in Table 11

Study	Study details	Findings
Mariscotti et al. (2017)	Retrospective analysis in which radiologists (n=2) blinded to histology reviewed s2DM and FFDM Participants: 231 Italian women with 250 biopsy-proven suspicious breast lesions (consecutive dataset) imaged with FFDM + DBT (combo mode) Unit: Hologic Selenia Dimensions unit with s2DM images reconstructed using C-view software	55/250 (22%) of lesions presented with microcalcifications <i>Lesion conspicuity where microcalcification was the mammographic feature</i> Seen better with s2DM reconstructed image: 48.1% Seen better with FFDM: 5.8% Equal visibility: 46.2%, $p = .0009$ <i>Concordance between s2DM and FFDM (BIRADS assessment)</i> Cohen's k (95% CI): 0.91 (95% CI: 0.8, 1) <i>Sensitivity for microcalcifications</i> s2DM: 92% FFDM: 85%, $p > .99$ <i>Specificity</i> S2DM: 43% FFDM: 43% $p > .99$ AUCs2DM: 0.64 ± 0.09 FFDM: 0.61 ± 0.09 , $p = .82$

Study	Study details	Findings
Berger et al. (2016)	Retrospective blinded reader (n=3) study Participants: 33 patients with total surgical resection with complete histological analysis proving DCIS Unit: Hologic Selenia Dimensions	<i>Correlation between DM and histology: Reader 1</i> FFDM: 0.879 DBT _{MLO/CC} : 0.914 <i>Correlation between DM and histology: Reader 2</i> FFDM: 0.717 DBT _{MLO/CC} : 0.783 <i>Correlation between DM and histology: Reader 3</i> FFDM: 0.735 DBT _{MLO/CC} : 0.770, $p < .001$
Clauser et al. (2016)	Retrospective blinded reader (n=4) study Participants: 150 FFDM and DBT cases (50 benign, 50 malignant, 50 BIRADS 1) from patients recalled to assessment or undergoing diagnostic workup Unit: Siemens Mammomat Inspirations and Hologic Selenia Dimensions	Detection range for 4 readers with FFDM: 89 – 96% Detection range for 4 readers with DBT alone: 92 – 97%, $p = .51$ No significant differences in visibility ratings <i>Sensitivity and AUC measurement</i> FFDM: 88-90%, 0.829 DBT alone: 78-98%, 0.833
Tagliafico et al. (2015b)	Multicentre reader (n=6) study Participants: 107 microcalcifications imaged with FFDM and DBT (66 malignancies, 41 benign lesions) Unit: Hologic Selenia Dimensions	<i>Sensitivity</i> FFDM: 100% DBT: 91.1% <i>Specificity:</i> FFDM: 94.6% DBT: 100%
Destounis et al. (2013)	Retrospective study Participants: 177 examinations with microcalcifications and recommended for needle core biopsy (50 malignancies, 116 benign) Unit: Hologic Selenia Dimensions	FFDM image quality superior to DBT: 7.8% DBT image quality equivalent to FFDM: 53% DBT image quality superior to FFDM: 39.2%

4.5. Masses

The shape, texture and appearance of tissue in and around a mass is indicative of whether it is benign or suspicious for malignancy: round or oval masses are often benign, spiculation is indicative of malignancy (with a PPV for malignancy as high as 97%) and a poorly defined mass with irregular contours is an early marker of cancer. Fifteen studies and four narrative literature reviews commented on masses as a mammographic presentation. Most studies describing masses did not differentiate between different types of mass (eg, branching, presenting with spiculation or oval, well-circumscribed masses). We have described where studies do provide differentiation. A summary of findings is included in *Table 14 (A-D)*.

Systematic reviews and narrative literature reviews

Systematic reviews: none

Four narrative literature reviews: Michell & Batohi (2018); Destounis et al. (2017); Mall et al. (2017); Peppard et al. (2015)

Pictorial atlas: Houssami et al. (2016)

RCTs and prospective studies

RCTs: none

Four prospective studies:

Women recalled to assessment: Caumo et al. (2018)

Only symptomatic women and/or women with a diagnosed breast cancer: Refaat & Matar (2015); Noroozian et al. (2012)

Diagnostic pathway unclear: Bahrs et al. (2017)

Retrospective studies

11 retrospective studies:

Women recalled to assessment: McDonald et al. (2017); Lourenco et al. (2014)

Women recalled to assessment and symptomatic women: Ray et al. (2015)

Only symptomatic women and/or women with a diagnosed breast cancer: Nakashima et al. (2017); Bian et al. (2016); Mariscotti et al. (2016); Raghu et al. (2016); Takamoto et al. (2013); Yang et al. (2013)

Diagnostic pathway unclear: Chan et al. (2017); Mariscotti et al. (2017);

Key findings

The shape, texture and appearance of tissue in and around a mass is indicative of whether a mass is benign or suspicious for malignancy. As noted in *section 4.3*, imaging with DBT often results in reclassification of primary mammographic finding from a focal asymmetry to a mass. Improved lesion conspicuity with DBT reduces tissue overlap making it easier to determine mass margin by removing 'noise' and making it easier to determine between masses suspicious for malignancy and those that are benign. There is now sufficient evidence that DBT provides superior performance in terms of improving readers' view of mass margins/soft tissue lesions.

4.5.1. DBT makes it easier to determine between masses suspicious for malignancy and benign masses compared to FFDM alone

Narrative literature reviews

Our search identified four narrative literature reviews (Michell & Batohi, 2018; Destounis et al., 2017; Mall et al., 2017; Peppard et al., 2015) and one pictorial atlas (Houssami et al., 2016) that presented high-level summaries about the role DBT plays in the detection of masses. The narrative reviews reported that DBT:

- reduces tissue overlap making it easier to determine mass margin by removing 'noise' and making it easier to determine between masses that are suspicious for malignancy and those that are benign, and
- can detect masses with spiculation smaller than 7mm and masses larger than 8mm and is superior to FFDM in for determining lesion size (see *section 3.5.5*).

Loss of edge characteristics in low-radiation dose images were raised as a potential issue with DBT especially if there is a similarity between mass density and the surrounding fibroglandular tissue (but this can be overcome using different DBT settings). Both Michell & Batohi (2018) and Destounis et al. (2017) reported that there is now sufficient evidence of DBT's superior performance in terms of conspicuity of soft-tissue lesions (based on studies by Seo et al., 2014, Amer et al., 2017. and Nakashima et al., 2017), which usually have a primary mammographic finding based on a mass.

Prospective studies of women recalled to assessment

Four prospective studies reported on the impact of DBT on masses as a mammographic finding. Two studies reported statistically significant improvements in mass classification with DBT compared to DM; one reported no difference. Caumo et al. (2018) completed a comparative analysis (n=314 cancers) using prospective data from the Verona screening program to describe the characteristics of cancers detected with DBT + s2DM compared to retrospective control data with FFDM screening. There was a small non-significant increase in masses or densities reported with s2DM compared to FFDM (66.4% of total cancers compared to 63.2%) and a borderline significant decrease in masses presenting with microcalcifications (8.9% of total cancers compared to 14.5%). The authors provided no further comments about the significance of this finding.

Prospective studies of symptomatic women recalled only

In a very small study by Refaat & Matar, DBT provided superior visibility in 55% of mass presentations and equivalent visibility in the remainder. In an earlier study, Noroozian et al. (2012) reported improved mass visibility with DBT compared to DSCV was explored by Noroozian et al. (2012). In their prospective study of 67 women with lesions classified as BIRADS 4 or 5, the authors reported that all four readers reported improved mass visibility with DBT compared to DSCV (with one reader's performance reaching statistical significance). A non-significant increase in AUC measurement favouring DBT was also reported.

Prospective studies: diagnostic pathway unknown

Bahrs et al. (2017) reported a statistically significant decrease of 6.6% in lesions classified as masses when DBT images were read ($p<.0001$). Given that the study population was women with lesions classified as BIRADS 3 at FFDM imaging, 7/21 asymmetries were down-graded to BIRADS 1 or 2 and two were upgraded to BIRADS 4, one of which on histology was a carcinoma no special type. No downgraded lesions had malignant findings at follow-up (term not described). While a small study, the authors concluded that DBT resulted in reduced follow-up of indeterminate lesions presenting as BIRADS 3.

Retrospective studies

The 11 retrospective studies reported consistent information about how masses are visualised with DBT compared to DM. Three studies reported on overall conspicuity of masses using DBT in a number of different ways. All studies reported that conspicuity increased significantly with the use of DBT:

- Yang et al. (2013) reported better visibility of asymmetry in more than 58% of masses with FFDM + DBT compared to FFDM
- In Chan et al.'s 2017 fully crossed reader study, conspicuity (rated by six radiologists with median experience of 20 years) was much higher for masses when seen on wide-angle DBT compared to DM (6.7 compared to 5.9; $p=.00001$), and
- Mariscotti et al. (2017) reported that s2DM had superior visibility compared to FFDM in 28.9% of cases and equal visibility in 60.3% of cases ($p=.0009$).

Improved conspicuity also led to changes in BIRADS classification with increased primary mammographic findings recorded as masses (and a commensurate decrease in asymmetry). Mariscotti et al. (2016) reported increased accuracy of classification compared to proven histology, with an 19% improvement with DBT. Mariscotti et al. (2017) also reported increased specificity with s2DM compared to FFDM (a non-significant increase of 3% for sensitivity and equivalent sensitivity). These results suggest that DBT or s2DM will identify more cancers presenting as masses and correctly dismiss more women without breast cancer compared to

FFDM. Chan et al.'s study, DBT improved the sensitivity of mass classification by about 16% with a corresponding decrease in specificity of 15%, indicating that more women with malignant asymmetries were correctly evaluated (i.e., as masses) with wide-angle DBT compared with the DM imaging protocol but that more false-positive classifications of BIRADS 4a and above were recorded (requiring unnecessary biopsy).

Other studies reported changes with the use of DBT that resulted in different case management. Using data from a screening population, Lourenco et al. (2014) reported that after the implementation of DBT, 827 lesions were recalled to assessment. Prior to implementation (1175 lesions), 8.9% of recalls to assessment were for masses. Implementation DBT resulted in a substantial increase in masses as a primary mammographic finding: 26.8% ($p<.0001$). Reasons for this were improved lesion margin visibility resulting in reclassification of asymmetry to masses. It is not clear whether these study result include a mix of prevalent and incident screening but given that data is drawn from a screening population, it is likely to be more representative of real-world clinical practice than some other studies reported in this literature review. Other studies also demonstrated increases in recall to assessment due to mass, but the findings were considerably smaller (for example, McDonald et al. reported an increase of approximately 5% and Raghu et al. reported an increase of 8%).

Bian et al. (2017) also reported specific findings about masses in Chinese women with dense breasts, reporting that DBT improved lesion margin visibility and that there was better visibility of spiculation (no specific data was reported however). AUC measurement, sensitivity and specificity for the detection of masses were all significantly increased with DBT in this study (see section 3.5.5). The authors concluded that improved conspicuity and 3D planes make it easier to differentiate between malignant and benign masses, a common refrain in the studies reporting on this mammographic finding.

Nakashima et al. (2017) described an emerging issue associated with circumscribed masses. As the authors note, on DM a circumscribed mass with a well-defined margin is most likely to be benign but they can also reflect malignancy. In Nakashima et al.'s study, two experienced radiologists blinded to clinical data and pathohistological findings reviewed both sets FFDM and DBT images to determine whether the mass was able to be detected and then rated visibility. For all masses, DBT was either equivalent (17%) or provided a superior view (83%). All sub-group analyses also reported equivalent or superior visibility but only results for breast density achieved statistical significance ($p=.016$). Nakashima et al. concluded that the reduction in overlapping breast tissue improved the conspicuity of circumscribed masses (especially in women with more dense breasts).

Table 14: Primary studies investigating the role of DBT in detecting breast cancers presenting as masses

14A Prospective studies of women recalled to assessment

NB: Study design described previously in Table 11

Study	Findings
Caumo et al. (2018)	<p><i>Cancers presenting as masses detected with:</i> DBT + s2DM: 209(66.4% of cancers detected with DBT+s2DM) FFDM: 96 (63.2% of all cancers detected with FFDM); $p=.5$ <i>Cancers presenting as masses with microcalcification detected with:</i> DBT + s2DM: 28(8.9% of cancers detected with DBT+s2DM) FFDM: 22 (14.5% of all cancers detected with FFDM); $p=.067$</p>

14B Prospective studies: diagnostic pathway unknown

NB: Study design described previously in Table 11

Study	Findings
Bahrs et al. (2017)	<p>Number of cases: Reader A</p> <p>FFDM: 19</p> <p>DBT: 12</p> <p>Number of cases: Reader B</p> <p>FFDM: 20</p> <p>DBT: 15</p> <p>Number of cases: Reader C</p> <p>FFDM: 23</p> <p>DBT: 18</p>

14C Prospective studies of symptomatic women only

NB: Study design described previously in Table 11

Study	Findings
Refaat & Matar (2015)	<p>Total malignancies presenting as masses: 20</p> <p>Equivalence rating based on reader's subjective view:</p> <p>DBT superior view compared to FFDM: 11 cases</p> <p>DBT equivalent view compared to FFDM: 9 cases</p> <p>DBT inferior view compared to FFDM: 0 cases</p>

14D Retrospective studies

NB: Study design described previously in Table 11

Study	Findings
Chan et al. (2017)	<p><i>BIRADS ≤ 3 malignancies</i></p> <p>FFDM: 30</p> <p>DBT: 14</p> <p><i>BIRADS ≤ 3 benign lesions</i></p> <p>FFDM: 222</p> <p>DBT: 207</p> <p><i>BIRADS ≥ 4a malignancies</i></p> <p>FFDM: 156</p> <p>DBT: 172</p> <p><i>BIRADS ≥ 4a benign lesions</i></p> <p>FFDM: 168</p> <p>DBT: 183</p>
Mariscotti et al. (2017)	<p>136/250 (54%) of lesions presented with masses</p> <p>Lesion conspicuity where mass was the mammographic feature</p> <p>Seen better with s2DM reconstructed image: 28.9%</p> <p>Seen better with FFDM: 10.7%</p> <p>Equal visibility: 60.3%, $p=.0009$</p> <p>Concordance between s2DM and FFDM (BIRADS assessment)</p> <p>Cohen's k (95% CI): 0.89 (95% CI: 0.67, 1)</p> <p>Sensitivity for mass</p> <p>s2DM: 94%</p> <p>FFDM: 94%, $p>.99$</p> <p>Specificity for AD</p> <p>s2DM: 75%</p> <p>FFDM: 78% $p=.5$</p> <p>AUC</p> <p>s2DM: 0.92 ± 0.02</p> <p>FFDM: 0.90 ± 0.03, $p=.56$</p>
McDonald et al. (2017)	<p>Change in recalled primary mammographic finding</p> <p>FFDM: 21.2%</p> <p>DBT: 26.9%</p> <p>Difference: 26.9 percent increase, $p<.001$</p> <p>Change in final BIRADS 3 finding/total findings</p> <p>FFDM: 15.7%</p> <p>DBT: 14.5%</p> <p>Difference: 7.6% decrease, $p=.674$</p>
Nakashima et al. (2017)	<p>DBT superior to FFDM</p> <p>Overall: 83%</p> <p>Low density: 68%</p> <p>High density: 91%, $p=.016$ for difference</p> <p>DBT equivalent to FFDM</p> <p>Overall: 17%</p> <p>Low density: 32%</p> <p>High density: 9%</p>

Study	Findings	
	between density	
Mariscotti et al. (2016)	Correct classification of mammographic findings FFDM: 70% FFDM + DBT: 89%, $p<.0001$	No further comment
Raghu et al. (2016)	Y0: 300/1374 (21.8%) lesions classified as BIRADS 3	Y3: 216/901 (24%) lesions classified as BIRADS 3
Ray et al. (2015)	19 lesions were FFDM-occult.	5/19 lesions were spiculated masses seen only on FFDM + DBT.
Lourenco et al. (2014)	Recalled to assessment with DM: 105 (8.9%).	Recalled to assessment with DBT: 222 (26.8%): mostly reclassified asymmetries, $p<.0001$
Takamoto et al. (2013)	37 breast cancers: Following DBT in affected breast, increase in BIRADS category for eight (no further data provided): OR=4.48, (95% CI: 1.33, 15.13; $p=.016$).	No further comment
Yang et al. (2013)	17 lesions presented as masses (28.8%) <i>Lesion conspicuity where microcalcification was the mammographic feature:</i> Seen somewhat better on FFDM + DBT: 11.8% Definitely seen better on FFDM + DBT: 47% Same: 41.2%	No further comment

5. CANCER CHARACTERISTICS

In cancer diagnosis, the gold standard test is pathological measurement. This may include examining markers like hormone receptor status, human epidermal growth factor receptor 2 (HER-2) status, and cell proliferation rate. Physical characteristics like tumour size, grade and lymph node status also play an important role ensuring accurate diagnosis of cancer type and determining treatment options and how an individual may respond to treatment/overall prognosis.

We want to know if there are any differences in the types of cancer detected with DBT compared to other imaging modalities (DM in particular). That is, does DBT detect more clinically significant breast cancers compared to DM and does it detect clinically significant breast cancers when they are smaller, and prognosis is likely to be better for women? Chapter 5 of this literature review describes evidence about differences in the specific cancer characteristics of malignancies identified with DBT compared to those identified with DM imaging. It includes:

- Cancer type
- Tumour stage and grade (including lymph node status), and
- Biomarker (receptor status).

The discussion includes a description of the number of studies identified, statements about the overall quality of the studies, and a summary of the results from all studies. Study tables provide additional detail about study population, methodology, intervention, comparator and key results. This information is used to answer Research Question 1 about the effectiveness of DBT in a breast cancer assessment clinic.

Key findings

DBT detects more invasive disease compared to DM

DBT, alone or in combination with FFDM, detects significantly more invasive cancers compared to FFDM. Pooled analysis drawn from studies of women participating in screening as well as women recalled to assessment reported this result consistently. Key results included:

- a relative risk (RR) of 1.327 for the increase of invasive cancer with FFDM + DBT compared to FFDM
- significant increases in detection of invasive ductal carcinoma (RR: 1.437) and special type carcinomas like tubular, papillary, medullary, and mucinous carcinoma (an increase of more than eight percentage points), with DBT (in a range of imaging protocols) compared to FFDM, and
- DBT-detected but DM-occult cancers are more likely to be invasive cancers.

Studies in other populations (including symptomatic women) reported the same result.

Mixed results are reported for the detection of invasive lobular carcinoma

ILC can be difficult to detect with mammography because a single cell or single files of cells may not result in clear changes at mammographic presentation. Pooled analysis from prospective studies set in screening populations reported increased detection of invasive lobular carcinoma (RR: 1.901). This finding is consistent with studies that reported increased conspicuity of subtle mammographic presentations like asymmetry and AD, which can be indicative of ILC. Two prospective studies reporting on ILC did not detect statistically significant increases in ILC

detection compared to FFDM. One retrospective study that specifically focused on ILC detection reported an increase in detection with DBT.

For women recalled to assessment or symptomatic women, DBT does not appear to detect more in situ carcinoma compared to DM

Pooled analysis results and results from prospective studies indicated that DBT did not appear to detect more in situ carcinoma including DCIS compared to DBT (RR 1.198, 95% CI: 0.942, 1.524). In the studies reported in this literature review, FFDM detected proportionally more DCIS cases across a range of studies in diagnostic populations (including populations of women recalled to assessment). DBT detected more cancers than FFDM overall but it did not preferentially detect more non-invasive cancers, which may never become clinically significant.

Invasive cancers may be detected at an earlier stage or different grade with DBT but there are some mixed results

DBT appears to result in more accurate determination of lesion margin and therefore results in more accurate sizing of tumours. Pooled analysis and prospective studies from screening populations indicates that DBT detects a higher proportion of early stage breast cancers compared to FFDM. For example, pooled analysis from studies set in screening populations indicated increased detection of T1 cancer (RR 1.388) or an increase in T1 and T1N0 cancers. T1N0 cancers are smaller and are likely to not have spread to auxiliary lymph nodes making them clinically important as these cancers detected at screening are more likely to have a better prognosis.

DBT's benefit in detecting cancers T2 or larger was less certain (RR 1.391, 95% CI: 0.895, 2.163). No consistent differences in detection was reported for Grade II or III cancers.

Mixed results were reported for DCIS staging, with data from the TOMMY trial and two other retrospective studies indicating that DBT detected larger DCIS with a higher grade, which supports some of the earlier findings that DBT may not accurately diagnosis cancers presenting as microcalcifications (although changes to reconstruction algorithms may have addressed this issue).

It is unclear whether DBT detects more node negative disease compared to FFDM

There were mixed results for the detection of node negative disease between the prospective and retrospective studies set in screening settings. The meta-analysis from screening/women recalled to assessment indicated an increase in the proportion of node negative disease detected with DBT, but other studies reported no difference in results by nodal status (or no change in grade or stage). This may reflect the screening imaging used (eg, FFDM + DBT or DCT + s2DM).

DBT and FFDM perform similarly in hormone receptor status

Regarding hormone receptor and HER-2 status of breast cancers, both modalities of (DBT and FFDM) imaging were comparable in their ability to detect both positive and negative status cancers.

5.1. Cancer type

Breast cancer tumours can be categorised into two categories:

1. Non-invasive cancers, and
2. Invasive cancers.

DCIS is a non-invasive breast cancer, located in the milk ducts. DCIS has not yet spread to surrounding tissue. While DCIS is non-invasive, if left untreated it has the potential to spread into surrounding healthy cells. If DCIS breaks through walls of its starting tissue and begins to spread into surrounding healthy cells it becomes an IDC but there is some uncertainty about which cases of DCIS will progress to invasive disease. LCIS is an area of abnormal cells and is considered to be a risk factor for the later development of invasive breast cancers. IDC is the most common breast cancer (around 80% of all breast cancers being this type) with ILC being the second most common breast cancer (Sharma et al., 2010). IDC tumours are usually defined by their irregular shape and hard texture. Subtypes of IDC (including invasive tubular, medullary, and papillary carcinomas) have differing cell arrangement, they generally all have a softer tumour texture. ILC tumours cells present in a uniform single file order and are a normal to firm texture.

Understanding the differing characteristics of types of tumours is fundamental in accurately and successfully using DBT and other imaging techniques as assessment and diagnosis tools. Increases in detection of invasive cancers is likely to be indicative of detection of more clinically relevant cancers which have poorer prognosis if not detected early. More DCIS cases may contribute to overdiagnosis if specific cases of disease were not going to progress to invasive cancers.

Findings from Allen + Clarke's 2018 literature review on DBT in screening: types of cancer detected

Allen + Clarke reported that DBT used as a primary screening test may result in increased detection of invasive cancer. Limited information about detection of DCIS was provided. While studies focused on diagnostic populations or cancer-enriched cohorts were excluded from the literature review on screening, some included studies reported on tumour characteristics. A fixed effect meta-analysis by Hodgson et al. (2016) reported data from the STORM and OTS trials finding a statistically significant increase in invasive cancer detection rate when DBT was used. Other primary studies supported the pooled analysis findings, reporting that FFDM + DBT resulted in the detection of more invasive cancers or smaller cancers but other primary studies reported no differences based on type, grade or size. Differences in mammographic presentation were reported (i.e., increased conspicuity for AD and stellate or spiculated masses and mixed results for microcalcifications). Tumour size at detection was reported to be slightly smaller with DBT in screening with a lower grade. Earlier studies showed that FFDM + DBT's performance did not appear to be superior for the detection of DCIS because of reduced visibility of microcalcifications. Later studies reported no differences in the types of cancers detected by either FFDM + DBT or FFDM.

Much of the evidence discussing cancer type/histology is included in literature from diagnostic populations. This literature review provides a more fulsome discussion of the characteristics of cancers identified with DBT imaging modalities compared to those identified with DM. Further information about differences in primary mammographic finding (including microcalcifications) is provided in *section 3.5* of this report.

Section 4.1 describes findings about the overall type of cancer identified with DBT compared to DM. Overall, the evidence base comparing type of cancers detected with DBT imaging protocols compared to DM imaging was surprisingly limited. While many studies described cancer type/histology in patient characteristics tables, few papers provided comparative analysis between different imaging modalities. Our literature search returned one systematic review, one narrative literature review and 14 studies that compared DBT combinations to FFDM. Stage and grade findings are reported in *section 4.2*. Primary studies already incorporated into systematic or literature reviews were reviewed but not separately assessed unless additional material not described in the systematic review or narrative literature review was included in the primary study.

Studies in this literature review are listed below.

Systematic reviews and meta-analyses

One systematic review: Yun et al. (2017)

One narrative literature review: Eghtedari et al. (2018)

RCTs and prospective studies

RCTs: none

Three prospective studies:

Only women recalled to assessment: Caumo et al. (2018)

Women recalled to assessment and symptomatic women: Chae et al. (2016)

Only symptomatic women and/or women with a diagnosed breast cancer: Mercier et al. (2015)

Retrospective studies

11 retrospective studies

Only women recalled to assessment: Bahl et al. (2018); Gilbert et al. (2015b); Nam et al. (2015); Wang et al. (2016)

Women presenting for screening and symptomatic women: Rafferty et al. (2014);

Only symptomatic women and/or women with a diagnosed breast cancer: Mariscotti et al. (2016);

Diagnostic pathway unclear: Mariscotti et al. (2017); Choi et al. (2016a); Elizalde et al. (2016); Thomassin-Naggara et al. (2015); Lång et al. (2014)

5.1.1. In screening populations, detection of invasive cancers increases with the use of DBT

This section describes detection of invasive cancers from studies set in screening populations (i.e., cancers detected in women recalled to assessment). This literature review reports only on the characteristics of cancers detected (rather results for overall cancer detection, which is a data more relevant to considerations about DBT as a screening tool).

Systematic review of breast cancers detected in a screening program

Our literature search identified one systematic review with accompanying meta-analysis carried out by Yun et al. (2017). This study investigated breast cancers detected in screening programs¹⁵ and we have included this review here given its overall relevance to the BSA program. Yun et al. compared cancer by type (invasive and carcinoma in situ) by imaging protocol, exploring histologic characteristics, tumour size at detection, and lymph node status. The main inclusion criteria were studies published prior to 31 December 2016 which compared the diagnostic value of FFDM + DBT and FFDM in a screening environment, with pathologic confirmation of lesions as the reference standard, cancer detection rate and pathologic characteristics. Studies were excluded if participants were not drawn from a general screening population, had study cohorts of fewer than 1000 women, were interpreted without blinding to reference standard results, or used data from other studies. QUADAS-2 was used to assess the methodological rigour of individual studies. Heterogeneity was not significant in any of the subset analyses. Yun et al. included data generated from 11 studies (four prospective trials and seven retrospective observational studies). Yun et al.’s analysis included 75,532 study participants who had FFDM + DBT and 175,825 participants who had FFDM. Further data from 37,085 study participants in fully paired trials related to both FFDM and FFDM + DBT was also included.

Pooled analysis found that FFDM + DBT increased the detection of invasive breast cancer compared to FFDM: the RR for invasive cancer detection was 1.327 (95% CI: 1.168, 1.508). The analysis also showed no increase in the detection of DCIS: RR 1.198 (95% CI: 0.942, 1.524). These findings suggest that while DBT detects more cancers overall, it detects clinically important cancers. Yun and colleagues also completed pooled analysis for a breakdown of cancer type providing relative risk estimates for IDC and ILC:

- IDC: RR 1.437(95% CI: 1.189, 1.737)
- ILC: RR 1.901 (95% CI: 1.213, 2.979).

Detection of both IDC and ILC detection improved with DBT + FFDM compared to FFDM in screening populations. No further analysis of other histological types of cancer due to a limited range of studies reporting on these.

Key findings
Increased detection of invasive cancer with FFDM + DBT (RR 1.327; 95% CI: 1.168, 1.508).
Benefit of FFDM + DBT was not increased for carcinoma in situ (RR 1.198; 95% CI 0.942, 1.524).
Increased detection of IDC and ILC with FFDM + DBT (RR 1.437, (95% CI 1.189, 1.737) (RR 1.901 95% CI 1.213, 2.979).

Prospective studies involving women recalled to assessment

DBT + s2DM compared to FFDM alone: detection of invasive cancers

One prospective study of women recalled to assessment only reported on overall detection of invasive cancers. Caumo et al. (2018) reported findings from a pilot evaluation of DBT + s2DM compared to FFDM. Participants were drawn from the Verona population-based breast cancer

¹⁵ Yun et al.’s study included data from the STORM trial, the Oslo trial and the Malmö trial as well as key American studies covered in *Allen + Clarke’s* literature review on DBT in screening (i.e., Powell et al., 2017; Sharpe et al., 2016; Durand et al., 2015; Lourenco et al., 2015; Greenburg et al., 2014; McCarthy et al., 2014; Rose et al., 2013).

screening program. During the DBT pilot period 34,071 women were screened (April 2015 – March 2017) and data compared to the 29,360 women examined during the previous screening period which used FFDM (April 2013 – March 2015). A total of 468 breast cancers were detected over both two-year periods (315 detected with DBT + s2DM; 153 with FFDM). Images from the two modalities were independently interpreted by two of four readers who performed readings during both the pilot study and during the FFDM screening periods. Caumo et al.'s results showed important and statistically significant differences in cancer type with DBT + s2DM compared to FFDM, with results consistent with those presented by Yun et al. The authors reported that compared to FFDM, DBT + s2DM:

- detected more special type invasive cancers (12.4% compared to 4.6%, $p=.008$), and
- identified proportionally fewer IDC (61.6% of cancers were IDC compared to 64.7% respectively, $p=.51$).

Non-significant results from this study showed that DBT + s2DM did not result in a significant increase in detection of ILC: DBT + s2DM identified slightly more ILC (14.0% compared to 12.4% with FFDM, $p=.65$) but an increase in ILC was expected given the improvement in lesion conspicuity of AD and asymmetry (described in *section 3.5.1* and *section 3.5.2*). Caumo et al. concluded that detection of additional special-type cancers represented finding more cancers that have good prognosis if detected early.

Retrospective studies of women recalled to assessment

Three retrospective studies of women recalled to assessment only reported on overall detection of invasive cancers with DBT (in a range of combinations) and DM. Results presented in these studies are broadly consistent with the findings from screening populations (reported in Yun et al.'s systematic review) and the prospective studies. That is, DBT alone or in combination with FFDM detects higher proportions of invasive breast cancers compared to FFDM alone.

FFDM + DBT compared to FFDM: detection of invasive cancers

In their retrospective analysis of screen-detected cancers and interval cancers, Bahl et al. (2018) retrospectively reviewed 78,385 FFDM and 76,896 FFDM + DBT screening images, for a total of 948 breast cancers, to determine if cancer rates and tumour characteristics differed between the two imaging protocols. Images were taken at two different timepoints: FFDM took place between January 2009 to February 2011 prior to the implementation of DBT; all FFDM + DBT screenings took place between January 2013 to February 2015. The authors found no significant differences between the two groups for patient age ($p=.45$), or breast density ($p=.75$). Breast cancers were considered screening-detected if diagnosed within 365 days of a positive screening exam. Interval cancers were those diagnosed within 365 days of a negative screening assessment. In line with results reported by Yun et al. (2018), Bahl et al. reported that FFDM + DBT resulted in detection of a higher proportion invasive breast cancers (74.2% with FFDM + DBT compared to 66.0% with FFDM, $p=.01$).

Two studies reported on cancers only detected on DBT (i.e., those that were FFDM-occult). Wang et al. (2016) retrospectively compared the pathology and histologic grading of breast cancers (comparing FFDM + DBT to FFDM). From a screening cohort of 12,444 women, 65 breast cancers were detected in 63 patients. Five readers independently reviewed each set of patient images, FFDM first and then FFDM + DBT. A breast cancer was considered occult if the no reader could detect it on the FFDM images, but it could be identified on in a DBT slice. Ten cancers were occult on FFDM, and nine of these were invasive cancers. The majority of DCIS found in all 65 breast cancers (96%) was identified by FFDM. While the sample of mammographically occult cancers investigated was very small, this study still highlights that additional cancers detected

on DBT are more likely to be invasive. increasing the invasive cancer CDR from 2.6 per 1000 screening examinations to 3.3 per 1000 screening examinations.

In a slightly larger study investigating FFDM-occult cancers, Nam et al. (2015) retrospectively investigated 41 cancers that were initially missed on DM and only detected by ultrasound. DBT was performed in combination with FFDM and the image sets were compared. Using DBT, readers were able to detect a higher proportion (54%) of missed cancers compared to FFDM (27%). Of the 41 cancers detected, 37 (90.2%) were invasive cancers and four (9.8%) were DCIS. When the area of concern was known for readers, an additional 29% of cancers were able to be detected. Whether the readers were told these areas, or whether they used the FFDM in addition to the DBT images to identify them is unclear.

5.1.2. Detection of invasive cancers increases in mixed (symptomatic/asymptomatic) populations or symptomatic populations only

The following studies suggest that DBT detects clinically relevant cancers in women with symptoms of breast cancer as well as asymptomatic women, and these results are consistent with the results from the studies set in screening populations. It is not clear if the DBT imaging modality has a significant impact given that the general direction of effect is similar for FFDM + DBT and other modalities such as DBT_{MLO} alone.

Prospective studies from mixed or symptomatic populations

The literature review returned two prospective studies comparing detection of invasive cancers. These studies include either asymptomatic and symptomatic women (Chae et al., 2016) or symptomatic women only (Mercier et al., 2015). Overall rates for invasive cancer detection were higher with DBT imaging. Increased ILC detection was not found in Mercier et al.'s study (as with Caumo et al.'s results) but increased detection was reported in Mariscotti et al.'s retrospective study and indicated in the increased conspicuity of AD and asymmetries.

Chae et al. (2016) reported on breast cancer type detected with DBT_{MLO} compared to FFDM in women with suspected breast cancer (BIRADS reading outcome of 4 or 5), or those with diagnosed breast cancer. No further information about the number of women in each category was provided in the article. All participants underwent both imaging techniques (DBT_{MLO} and FFDM) and both sets of images were reviewed by three experienced readers. After reviewing 337 breast cancers, subgroup analysis showed DBT_{MLO} had a higher sensitivity than FFDM for detecting invasive cancers (DBT_{MLO} was 90.6% compared to 83.4% for FFDM; $p=.004$).

Mercier et al. (2015) found that while DBT improved sensitivity of detection for ductal types of cancer (no data provided but $p=.016$) compared to FFDM, the same improvement was not seen in lobular disease (no data provided but $p=.17$). This study was much smaller than Caumo et al.'s with only 75 patients (74 women and one man) who had presented with approximately 124 lesions. These participants were an enriched sample as patients were only recruited if they presented with a BIRADS 4 or 5 lesions, therefore our ability to extrapolate these findings for the general screening population is limited.

Retrospective studies from mixed or symptomatic populations

FFDM + DBT compared to FFDM: detection of invasive cancer

Two different methods of using combined DBT were investigated by Rafferty et al. (2014). Fifteen radiologists retrospectively read 310 cases sequentially using FFDM then added one view of DBT at a time.

For the detection of invasive cancers, sensitivity was:

- 60.6% for FFDM
- 72.6% for FFDM + 1v-DBT, and
- 82.3% for FFDM + DBT.

Statistical significance was achieved for the reported increase between FFDM and FFDM + 1v-DBT (12%, $p < .001$) and FFDM compared to FFDM + DBT (21.7%, $p < .001$).

As noted by Eghtedari et al. (2018), ILC can be difficult to detect with mammography because a single cell or single files of cells may not result in clear changes at mammographic presentation (i.e., ILC is more likely to present as a subtle mammographic finding like asymmetry or AD). It is also the second most common breast cancer type. ILC often presents as long single cell rows that diffuse invasively into tissues which allow them to grow without the formulation of a tumour mass. DBT has been shown to be more effective in the detection of massless distortions such as AD compared to FFDM (see *section 3.5.1*). Mariscotti et al. (2016) compared FFDM and FFDM + DBT in 83 patients with diagnosed ILC, finding that ROC analysis showed an AUC measurement of 0.89 (95% CI: 0.88-0.91) for DBT which was significantly higher than the AUC for DM alone (0.84, 95% CI: 0.82-0.86). The addition of DBT to FFDM also significantly increased the pooled sensitivity from 0.70 to 0.85 and significantly improved the detection of ILC subtypes: including classic (83% compared to 68%), solid (79% compared to 63%), and mixed (92% compared to 77%). No significant differences were seen between the two modalities for alveolar and pleomorphic subtypes. These results align with Yun et al.'s findings but present different results from the prospective studies (which could be due to having a larger sample size of ILC).

s2DM compared to FFDM: detection of invasive cancers

In 2017, Mariscotti et al. published another article on a larger retrospective study (148 breast cancers) comparing s2DM to FFDM. In this study, s2DM detected fewer false negatives for ILC (2 less FN), IDC (1 less FN), and "other invasive breast cancers" (unspecified) (3 less FN) and DCIS (1 less FN). Due to small numbers, none of these differences were significant (for example, 12/148 FN for s2DM compared to 19/148 FN for FFDM).

Other DBT combinations compared to FFDM: detection of invasive cancers

Three studies reported on different combinations of DBT alone compared to FFDM and the type of cancers detected with both.

- Thomassin-Naggara et al. (2015) found that FFDM + DBT_{MLO} can detect up to 26.2% of cancers that are FFDM occult. All lesions in their study that met this definition were found to be invasive cancers. This study was carried out in 150 women who underwent both FFDM and DBT_{MLO}.
- In a study with 84 breast cancers, Elizalde et al. (2016) reported DBT_{MLO/CC} detected an additional 17.8% invasive cancers. No significance tests were reported for these results.
- Lång et al. (2014) analysed discrepant breast cancer detection comparing DBT_{MLO} and FFDM and found a significant increase in the proportion of ILCs detected with DBT. The authors reviewed 26 discrepant cases (19 DBT-only detected and 7 DM-only detected). These ILCs presented mostly as spiculated masses under DBT imaging. The significant increase in ILC observed was only seen when the DBT ILC findings were compared with the proportion of ILC from their previous detection study.

5.1.3. In screening populations, detection of DCIS does not appear to increase if DBT is used

Systematic review of breast cancers detected in a screening program

Yun et al.'s 2017 systematic review of data on cancer type in screening populations reported mixed results regarding FFDM + DBT's ability to detect DCIS compared to FFDM. The authors noted that some studies suggested that DBT was inferior in the detection of microcalcifications (and therefore DCIS). Yun et al. (2017) highlighted results from two studies set in screening populations:

- Friedewald et al. (cited in Yun et al., 2017) reported no increase in detection of DCIS, and
- Conant et al. (cited in Yun et al., 2017) reported that FFDM + DBT had a higher detection rate for DCIS than FFDM.

Yun et al. (2017) hypothesised that as DCIS commonly presents as microcalcifications and not masses or distortions, DBT may be less suited for the detection of DCIS. Detection of microcalcifications is further discussed in *section 3.5.3*.

Prospective studies of women recalled to assessment

One prospective study of women recalled to assessment only reported on overall detection of DCIS. Caumo et al. (2018) (methodology described in *section 5.1.1*) reported that results from a pilot evaluation of DBT + s2DM compared to FFDM. In this study, DBT + s2DM detected fewer DCIS:

- A lower proportion of Tis (i.e., DCIS) cancers (11.4% vs 18.8%, $p=.037$), and
- A lower proportion of stage 0 (DCIS) cancers; (10.8% vs 19.2%, $p=.013$).

Non-significant results regarding an increase in detection of DCIS with micro-invasion were also reported: 0 tumours with FFDM; 0.6% of tumours identified with DBT + s2DM, ($p=1$, very small sample size). In this study, DCIS made up a lower proportion of all types of cancer detected with DBT + s2DM and Caumo et al.'s work adds further to the growing evidence base about the use of DBT and its contribution to overdiagnosis through the detection of additional DCIS cases.

More information about the size of tumour at detection and the detection of Tis cancers from this study is included in *section 5.2.1*.

Retrospective studies of women recalled to assessment

Two retrospective studies of women recalled to assessment only reported on overall detection of DCIS with DBT (in a range of combinations) and DM. Results presented in these studies are broadly consistent with the findings from screening populations (reported in Yun et al.'s systematic review) and the prospective studies. That is, DBT alone or in combination with FFDM detects slightly lower proportions of DCIS compared to FFDM alone.

FFDM + DBT compared to FFDM: detection of DCIS

Bahl et al. (2018) (methodology described in *section 4.1.1*) reported that the use of FFDM + DBT resulted in detection of a lower proportion of in situ cancers (25.8% with FFDM + DBT compared to 34% with FFDM, $p=.01$). The authors concluded that FFDM + DBT did not increase DCIS but instead detected more invasive cancers. No further detailed description of histology was reported in this study.

Gilbert et al.'s (2015b) Health Technology Assessment reporting on TOMMY trial data (described in *section 3.1*) did not provide a breakdown by cancer type. Instead they provided

some overarching data which is consistent with the systematic reviews and prospective study data. Reporting on matched comparison data (all images read three ways) for 1112 cancers, Gilbert et al. (2015b) reported detection of DCIS cases by reading arm, finding that detection of DCIS was not increased with DBT, which is consistent with other study results:

- DCIS detected by FFDM: 174 cases
- DCIS detected by FFDM + DBT: 174 cases, and
- DCIS detected by DBT + s2DM: 169 cases.

No significance testing for these results was provided.

5.1.4. In mixed study populations, evidence suggests that DBT does not increase DCIS detection compared to DM

The following studies suggest that DBT does not detect more DCIS in women with symptoms of breast cancer as well as asymptomatic women compared to FFDM. This suggests that DBT may not contribute as significantly to overdiagnosis as feared; however, there are some mixed (but not significant) findings.

Prospective studies in mixed study populations

In Chae et al.'s study (methodology described in *section 4.1.3*), no difference was reported for DCIS detection (DBT_{MLO} was 70% for compared to 53.3% for FFDM; $p=.227$). While in a mixed study population of asymptomatic women and those with symptoms, this result is consistent with the findings presented in *section 5.1.3* (DCIS detection in women recalled to assessment).

Retrospective studies in mixed study populations

FFDM + DBT compared to FFDM alone: detection of DCIS

In Rafferty et al.'s (2014) study (methodology discussed in *section 4.1.3*), sensitivity for in situ cancers rose slightly with the addition of DBT views (but significance for the differences was not achieved):

- 67.5% for FFDM
- 68.8% for FFDM + 1v-DBT, and
- 70.8% for FFDM + DBT).

This suggests that, like other studies, DBT does not result in a large increase in DCIS.

s2DM compared to FFDM: detection of DCIS

In Choi et al.'s (2016a) retrospective observer study, interpretive performance of s2DM (reconstructed from DBT) was compared to FFDM. In this study, 107 T1 invasive breast cancers were detected (214 patients). Of these, 93 were non-calcified lesions (not further described) and 14 were calcified. No results achieved statistical significance and overall performance between the two methodologies was very similar for the three readers (eg, using s2DM, three readers detected between 58.1% and 66.7% non-calcified cancers compared to 57.0% and 66.7% with FFDM). Comparability between detection of calcified cancers was also equivalent, an important finding given previous concern about the ability of DBT to accurately detect microcalcifications compared to FFDM (i.e., s2DM does not underestimate cancers presented with microcalcifications).

Other DBT combinations compared to FFDM: detection of DCIS

In a study with 84 breast cancers, Elizalde et al. (2016) reported DBT_{MLO/CC} detected no additional DCIS compared to FFDM alone. No significance tests were reported for these results.

5.2. Tumour stage and grade

Breast-screening with FFDM has demonstrated its effectiveness in reducing cancer mortality by detecting breast cancers when they are smaller, node negative and of a lower stage and grade (all of which are associated with improved patient outcomes). The assessment of newly diagnosed breast cancer is essential to obtain an estimate of staging, which is integral in prognosis development. Staging is used to describe the characteristics of the cancer. It involves determining the extent of disease in the affected breast and in the contralateral breast, evaluating regional lymph nodes and identifying other sites of disease if the cancer has metastasised. Staging is also used to assist in treatment planning and informs follow-up surveillance. The TNM classification along with other measures are used to determine a patient's overall cancer stage (stage 0 – stage IV). Breast cancers are staged by the TNM system: primary tumour size (T), whether regional nodes are involved (N), and distant metastasis (M). Each section is assigned to a subcategory depending on the lesion's characteristics (such as size, margin outline, and extent of growth). Imaging and histological and pathology testing are used to place breast cancers into the correct stage and subcategory.

Breast cancer grading is conducted histologically after a successful biopsy is completed and depends on how the tumour cells differ from healthy cells. Grade 1 breast cancer cells look small and uniform like healthy cells and are usually slow growing. Grade 3 breast cancer cells appear abnormal, usually due to a much faster rate of growth.

Early detection of breast cancer is one of the key ways in which mortality is reduced (i.e., smaller cancers have better prognosis). We want to know whether DBT results in differences in the stage and grade of cancers detected compared to those detected with DM. In this literature review, one systematic review, three literature reviews and 11 studies reported comparative results on the stage and grade of cancers. Primary studies already incorporated into systematic or literature reviews were reviewed but not separately assessed unless additional material not described in the systematic review or narrative literature review was included in the primary study.

The studies discussed in this literature review are listed below.

Systematic and/or literature reviews

One review: Yun et al. (2018)

Three reviews: Destounis et al. (2017); Moseley et al. (2016); Peppard et al. (2015)

RCTs and prospective studies

RCTs: none

Five prospective studies:

Only women recalled to assessment: Caumo et al. (2018)

Women recalled to assessment and symptomatic women: Chae et al. (2016); Seo et al. (2016)

Only symptomatic women and/or women with a diagnosed breast cancer: Mercier et al. (2015)

Diagnostic pathway unclear: Helal et al. (2017)

Retrospective studies

Six retrospective studies:

Only women recalled to assessment: Bahl et al. (2018); Choi et al. (2016a); Wang et al. (2016); Gilbert et al. (2015a)

Diagnostic pathway unclear: Berger et al. (2016); Freer et al. (2015)

5.2.1. In women recalled to assessment, there were mixed results on whether DBT (either with FFDM or s2DM) enables better assessment of tumour margins (and therefore size) and detects cancers at an earlier stage and grade compared to FFDM alone

Systematic and narrative literature reviews

Our literature search identified one systematic review with accompanying meta-analysis carried out by Yun et al. (2017). The methodology for this review is described in *section 5.1.1*. Yun et al.'s pooled analysis of invasive cancer detection rates by T-stage included five studies (70,985 participants). In women recalled to assessment, Yun et al. (2017) reported that, compared to FFDM, FFDM + DBT had a significantly higher detection rate for Grade I and II/III breast cancers, and T1 cancer (pooled RR of 1.388) (see box, below). FFDM + DBT also resulted in an increase in the detection of early invasive breast cancer (stage T1N0). No difference between FFDM + DBT and FFDM was seen in the detection of grade III cancers. The added benefit of DBT for improved T2 cancer detection was not proven. Together, these findings suggest that DBT is useful for the detection of early stage invasive cancers, allowing women to begin treatment earlier.

Key findings Yun et al. (2017)

Increased detection of invasive cancer at T1 stage ($\leq 2\text{cm}$) (RR 1.388, 95% CI: 1.137, 1.695)
Uncertainty in the detection of T2 or larger cancers (RR 1.391, 95% CI: 0.895, 2.163).
Increased detection of grade I (RR 1.812, 95% CI: 1.372, 2.393)

Narrative literature reviews by Destounis et al. (2017), Moseley et al. (2016), and Peppard et al. (2015) did not specifically report on TNM staging of breast cancer; however, they described studies of DBT's ability to assess lesion size and margin outline (including Yun et al.'s meta-analysis). Key study results are included in *Table 15* (below). Included studies reported that DBT appears to be at least comparable to (and often superior to) FFDM in the assessment of breast cancers, with increased concordance with pathology and generally more accurate lesion sizing. Taken together, these results mean that DBT is likely to improve the grading and staging of breast cancers, but further evidence of this is required in populations of women recalled to assessment as it is not clear from the narrative reviews which diagnostic pathways were included.

Table 15: Study summaries from three narrative literature reviews

Study	Design	Results
Amer et al. (2017)	Retrospective study of 102 patients	Accuracy for determining lesion margin was superior with DBT (69%) compared to with FFDM (40%).
Luparia et al. (2013)	Retrospective study reviewing 149 breast cancers in 110 women	DBT had a 66.4% concordance with pathology compared to 54.4% with FFDM. Additionally, over-estimation of tumour size was lower in DBT (11.4%) compared to FFDM (14.1%).
Mun et al.	Retrospective image	FFDM mis-sized 50 lesions by more than 1cm compared to 33 mis-sized with

Study	Design	Results
(2013)	analysis of 173 breast cancers	DBT ($p=.003$). Sizes were over-estimated in 39 cases and under-estimated in 11 with FFDM compared to 25 and 8 with DBT. Standard error was 3.5% compared to 3.0%.
Skaane et al. (2013)	Oslo Tomosynthesis in Screening Trial	More invasive cancers detected, and most are node negative
Svane et al. (2011)	Retrospective study of 144 suspicious lesions	DBT had higher quality diagnostic images compared to DM but there was comparable overall accuracy.
Förnvik et al. (2010)	Prospective study of 62 women with 73 breast cancers	Tumour margin was determined in more cases. Staging was more accurate with DBT in tumours up to 20mm (63 cases) compared to FFDM (49 cases) ($p<.05$). DBT + ultrasound and FFDM + DBT correlated well with pathology ($R=0.86$ and $R=0.85$). Both were better than FFDM ($R=.071$). Radiologists measure margins more accurately.

Prospective studies in women recalled to assessment

Caumo et al. (2018) (methods described previously) found that DBT + s2DM was significantly more accurate in assessing tumour stage (including tumour size) compared to FFDM. DBT + s2DM detected a significantly:

- higher proportion of T1a cancers (17.0% vs 7.6%, $p=.008$), but no significant difference in detection ability in the two methods was observed for T1b and T1c breast cancers
- lower proportion of T2 cancers; (8.3% vs 15.3%, $p=.027$), and
- higher proportion of stage IA cancers (65.0% vs 53.6%, $p=.019$); stage IB cancers (8.5% vs 1.3%, $p=.003$); and stage IIA cancers (8.2% vs 16.6%, $p=.007$).

Caumo et al. found that DBT + s2DM detected a significantly higher proportion of grade I ($p=.003$) and grade II cancers ($p=.05$) compared to FFDM. No significant difference in detection ability in the two methods was observed for stage IIB, III, and IV breast cancers. These findings show that while DBT identifies more cancers, it appears to not be preferentially detecting DCIS, lessening concern surrounding the possibility of it increasing the harmful effects of overdiagnosis.

Retrospective studies in women recalled to assessment

Four retrospective studies (including data from the TOMMY trial) reported on tumour stage and grading with some mixed results reported regarding the stage and grade of malignant DCIS. While retrospective literature findings are mixed, higher quality studies suggest that FFDM + DBT detects more lower grade (I and II) invasive cancers compared to FFDM.

FFDM + DBT compared to FFDM: tumour staging and grading

Four studies reported on FFDM compared to FFDM + DBT, with mixed results. Bahl et al. (2018) (study described in *section 5.1.1*) reported no difference between FFDM and FFDM + DBT for:

- the size of invasive cancers detected at screening (1.2cm compared to 1.1cm, $p=.27$).
- the proportion of invasive cancers detected at Grades 1, 2, and 3 (which were similar across both methodologies).

In Bahl et al.'s study, in situ cancers were slightly larger when detected with FFDM + DBT (1.3cm compared to 1.1cm with FFDM, $p=.01$) and more were Grade 2, but fewer were Grade 3 compared to those detected with FFDM alone. Bahl et al. (2018) also reported on interval

cancers (n=83). For these, FFDM + DBT detected slightly smaller invasive cancers (1.6cm compared to 2.0cm) but with a similar grade distribution. Interval in situ cancers detected with DBT (n=17) were larger (1.5cm compared to 0.8cm, $p=0.048$) and more likely to be of a higher grade than those detected with FFDM. These results differ from those presented in the systematic review and in other studies, which may reflect the impact of using the combined imaging modality (FFDM + DBT).

Like Bahl et al., Wang et al. (2016) reported that there were no differences in the size of invasive cancers identified with FFDM + DBT compared to FFDM. While FFDM-occult lesions were smaller (median: 6 mm, range 3-13 mm) than those identified by FFDM (median: 10 mm, range 4-43), significance was not achieved (probably due to a small sample size). Wang et al. reported that breast cancers that were FFDM-occult were significantly associated with a grade 1 histological score ($p=0.02$). Detection rates were similar for grade II and III cancers, which is a similar finding for both Bahl et al. and Gilbert et al. (2015a).

s2DM compared to FFDM: Tumour staging and grading

Results from the TOMMY trial reported that FFDM + DBT and DBT + s2DM were comparable at detecting most cancer types and characteristics; however, Gilbert et al. (2015a) reported that:

- DBT + s2DM was inferior for depicting microcalcifications and 11-20 mm DCIS compared to FFDM (but FFDM + DBT had a similar performance)
- DBT + s2DM and FFDM + DBT had a higher sensitivity for detecting 11-20 mm invasive breast cancers compared to FFDM, and
- FFDM + DBT had a significantly higher sensitivity in detecting grade II invasive cancers compared to FFDM (91% [95% CI: 88, 94], vs 87%; [95% CI: 84, 90], respectively, $p=0.01$).

In a single-centre retrospective study, Choi et al. (2016a) found that s2DM was comparable to FFDM in detecting T1 stage breast cancers, suggesting that FFDM + DBT may be unnecessary if DBT + s2DM is used. In this study, 107 patients with unilateral invasive breast cancer ≤ 2 cm (T1-stage), no clinical symptoms before diagnosis, and who had not had neoadjuvant chemotherapy prior to surgery were matched to 107 negative cancer results. Choi et al. found that s2DM was comparable to FFDM for detecting and visualising T1-stage breast cancers, suggesting that s2DM may be an acceptable alternate method for the detection and characterisation of early invasive cancers. This would result in faster imaging sessions and a significant reduction in radiation dose, and may avoid some of the issues associated with the staging/grading of DCIS detected with DBT as discussed in Bahl et al. These findings support early concerns discussed the literature that DBT alone may miss clinically significant DCIS, but that performance is improved when DBT + s2DM is used.

5.2.2. In studies with mixed populations (symptomatic/asymptomatic women), there were mixed results on whether DBT (either with FFDM or s2DM) enables better assessment of tumour margins (and therefore size) and detects cancers at an earlier stage and grade compared to FFDM alone

Prospective studies

Four other prospective studies have shown similar T-staging results (i.e., the detection of smaller cancers).

Mercier et al. (2015) (75 breast cancers imaged with DBT) showed that DBT had a consistent sensitivity (68.5%) irrespective of the size of tumours and was superior to FFDM for lesions

under 1 cm³ ($p=.004$). Chae et al. (2016) also reported DBT_{MLO} as having a significantly higher sensitivity for detecting T1 stage breast cancers over FFDM and reported significantly higher sensitivity in the detection of stage T2-3 lesions (>2cm). This study included 337 malignant cancers. Helal et al.'s (2017) study of 98 women with proven malignant breast cancer investigated the T-staging accuracy of three imaging modalities: FFDM, DBT and contrast-enhanced spectral mammography (CEM). DBT correctly identified the size of more masses than FFDM (69 compared to 59). DBT was superior to FFDM in T-staging but was only slightly more effective at detecting early stage cancers compared to FFDM (detecting an additional three cancers). No statistical testing was completed for this study, so the validity of the result is uncertain (but is consistent with other study findings).

DBT's superiority compared to FFDM in the size measurement of breast lesions was also demonstrated in a preoperative study population. Seo et al. (2016) also analysed lesion size, reporting that DBT is likely to be superior in measuring lesion, with more readers correctly identifying larger lesions. In their study, almost all readers detected lesions of approximately 2.9 cm; only few readers (1-2) detected lesions of around 1.6 cm. These results were only for FFDM +DBT and unfortunately the same data was not presented for FFDM, so detection performance could not be compared.

Retrospective studies

Two studies (diagnostic pathway unclear) reported on tumour grade and stage. Freer et al. found the mean size of invasive cancers detected to be 7.5 mm. Unfortunately, no data was available for on the tumour sizes that were visible on DM therefore no true comparisons could be made. However, when compared to results from other studies it provides some evidence suggesting that DBT-only detect lesions are of clinical significance.

Berger et al. (2016) also looked at the size of DCIS detected by DBT compared to FFDM. In their study of 33 DCIS cases, the authors reported that the correlation between DBT and histology was higher than the correlation between FFDM and histology for each of the three readers. The correlation between DBT and histology was 0.914 ($p=.01$); correlation between FFDM and histology was 0.879 ($p=.01$). Berger et al. found that lesions were mis-sized less with DBT (15.2% \pm 9.1%) compared to FFDM (16.5% \pm 11.5%). This led the authors to conclude that, while both FFDM and DBT under-estimated the size of DCIS, DBT was slightly better at describing the extent of DCIS but no statistical testing was completed.

5.2.3. In women recalled to assessment, there were mixed results for DBT's detection of invasive cancer with nodal metastasis

Five studies reported on differences between FFDM and DBT in terms of nodal status. All these studies involved women recalled to assessment. Compared to FFDM, DBT appears to increase the detection of early stage breast cancers that have not yet spread to surrounding auxiliary lymph nodes but there are some mixed results from large studies reporting on this issue.

Systematic review

Yun et al. (2017) (methodology described in *section 4.1.1*) also looked at six studies (78,395 participants), which evaluated FFDM + DBT performance in detecting the presence of nodal metastasis in invasive breast cancers detected in women recalled to assessment. FFDM + DBT detected more invasive cancer without nodal metastasis, but the pooled RR for invasive cancer with nodal metastasis showed no improvement compared to FFDM (see box, overleaf). FFDM + DBT resulted in a significant increase in the detection of node-negative cancers (RR 1.451) showing that FFDM + DBT detects significantly more stage N0 invasive cancers compared to FFDM; however, there was no significant difference between the two imaging methods for stage

N1 and above. Detecting a higher proportion of small, node negative (Stage T1-2, N0) cancers is preferential to screening programs as patients with this diagnosis at screening has been shown to correlate with better prognostic outcomes (Neville et al., 1992).

Prospective studies

Smaller studies did not reflect the results of Yun et al.'s meta-analysis.

Caumo et al. (2018) reported no significant difference between DBT + s2DM and FFDM across all nodal stages (that is, DBT + s2DM neither detects cancer with more or less nodal involvement). Results included that 83.8% of DBT + s2DM-detected cancers were N0 compared to 80.5% with FFDM, 17.8% were N1 with DBT + s2DM compared to 13.6% with FFDM, and 1.7% compared with 2.7% with FFDM. This imaging protocol differs from the studies included in the systematic review, which may account for the difference in overall finding.

Retrospective studies

Three retrospective studies presented similar results to those reported by Caumo et al. (2018): that is, DBT when used as an adjunct to FFDM does not influence nodal status at detection of cancer.

- Bahl et al. (2018) reported no significant difference in detection rates of the nodal-status of invasive breast cancers between FFDM and FFDM + DBT (11.1% with DBT compared to 14.2% with FFDM; $p=.28$) (which differs from the systematic review findings)
- Results from the TOMMY trial showed that compared to FFDM, FFDM + DBT had a slightly higher sensitivity for detecting node-negative invasive cancers (88% vs 90%, respectively), those with 1-3 lymph nodes (86% vs 88%, respectively) and more than three lymph nodes (86% vs 93%, respectively) with cancer cells present (but none of the results were significant)
- Wang et al.'s reported positive nodal status only in cancers detected by FFDM, but this finding was not found to be significant ($p=.6$) which was most likely due to the low numbers of occult cancers identified in the study.

5.3. Biomarkers (receptor status)

Breast tissue growth is controlled by hormones. Approximately 70% of breast cancers are sensitive to specific hormones, meaning that presence of these cells in a breast cancer may result in faster proliferation of cancerous cells.

There are three main hormones/gene expressions:

1. oestrogen receptor positive/negative (ER+/-)
2. progesterone receptor positive/negative (PR+/-), and
3. human epidermal growth factor receptor 2 (HER-2).

A cancer may test positive for any one of these, all of them or none (triple negative disease). All breast cancers are tested for receptor status which then informs treatment planning, hormone therapy and prognosis. HER-2 cancers are more likely to be aggressive. ER+ and/or PR+ cancers are commonly associated with masses presenting with an irregular shape and are spiculated whereas HER-2 cancers are associated with microcalcifications with a branching morphology.

Three papers reported on receptor status. All were study populations of women recalled to assessment. Studies discussed in this literature review are below.

Systematic and/or literature reviews

No systematic reviews or narrative literature reviews

RCTs and prospective studies

RCTs: none

One prospective study: Caumo et al. (2018)

Retrospective studies

Two retrospective studies: Bahl et al. (2018); Lee et al. (2017)

Fewer studies have provided an analysis by receptor status but all report a consistent finding that there was no difference in receptor status of cancers detected with DBT imaging or FFDM in populations of women recalled to assessment. Further studies are needed to confirm these findings. The results are:

- In Caumo et al.'s prospective study, no difference was found in the detection rates of oestrogen receptor status, progesterone receptor status ($p=.75$) nor HER-2 status; however, data was missing from some of the HER-2 analysis.
- Bahl et al (2018) reported no significant difference in hormone receptor status in screening-detected invasive cancers between DBT and DM but the study was underpowered to carry out the same analysis in cases of DCIS.
- Lee et al. (2017) also found similar results in their retrospective multi-reader study of 288 cancers imaged with both FFDM and DBT. Of the cancers detected 194 (67%) were hormone receptor positive, 48 (17%) were HER-2 positive, and 46 (16%) were triple negative breast cancers (both hormone receptor negative and HER-2 negative). Lee et al. found that the differing tumour characteristics of molecular subtypes were able to be identified on DBT imaging, but these characteristics did not affect cancer detectability.

6. DBT’S ROLE IN BIOPSY

Biopsy of a suspicious lesion is often performed as part of the work-up if mammographic or ultrasound imaging shows an abnormality suspicious for breast cancer. There are three commonly used types of biopsy:

- 1. Fine-needle aspiration (FNA)
- 2. Core biopsy (including vacuum-assisted core biopsy (VAB), and
- 3. Surgical biopsy.

Types of Biopsy
<p>Stereotactic needle biopsy is a biopsy guided by mammography.</p> <p>FNA uses a hollow needle attached to a syringe to withdraw a small amount of tissue or fluid from a suspicious lesion. It is often used as a less invasive method for determining between a fluid-filled cyst or a solid mass. If a solid mass is found, VAB is done.</p> <p>VAB requires a small incision to be made where the needle will be inserted. Several tissue samples are taken using a vacuum-powered device.</p> <p>Surgical biopsy occurs under local anaesthetic. A three-to five-cm incision is made and tissue from the suspicious lesion and some healthy tissue (margin) is removed. Occasionally, if the lesion cannot be easily palpated, preoperative image-guide wire localisation is used to help the surgeon locate the lesion.</p> <p>Core biopsy is a biopsy completed by extracting a small tissue sample with a needle.</p>

Choice of biopsy type depends on the characteristics of the abnormality, with core-biopsy and surgical biopsy being the most frequently implemented. DM is frequently used during biopsy procedures to assist in guiding the biopsy needle to ensure that the tissue sample collected contains tissue from the area(s) of interest. DBT is increasingly used to perform guided biopsies of targets that are not clearly seen on DM (such as subtle AD or asymmetry) or ultrasound (Eghtedari et al., 2018). DBT is also beginning to be used to assist in the preoperative localisation of tumours; however, the available evidence base only discusses DBT-guided VAB at this stage.

Research Question 2 in our literature review is “Does DBT-guided biopsy improve lesion localisation compared to biopsy guided by DM or ultrasound imaging?” and “Does DBT-guided biopsy result in fewer procedural complications compared to biopsy guided by digital mammography or ultrasound imaging?” This literature review describes findings on the performance of DBT in lesion localisation and in the image-guided biopsy of lesions.

A total of 12 articles discussed DBT’s role in guiding biopsy. All study populations were women undergoing biopsy (diagnostic pathway unknown unknown) or women with diagnosed breast cancer. These articles discussed DBT-guided VAB, either alone or compared stereotactic VAB. No pooled analysis of results relating to DBT-guided biopsy was identified. Five narrative literature reviews provided limited discussion of DBT-guided biopsy findings. Most related to the overall performance of DBT-guided VAB; however, one prospective study only assessed patient experience.

No primary studies clearly identified or discussed a reduction in procedural complications with the use of DBT-guided VAB although this may change as the evidence base develops further.

Study summaries are included in *Table 16*. The studies discussed in this literature review are listed overleaf.

Narrative literature reviews

Five reviews: Eghtedari et al. (2018); Michell & Batohi (2018); Destounis (2017); Destounis et al. (2017); Mall et al. (2017)

RCTs and prospective studies

RCTs: none

Two articles:

Diagnostic pathway unclear: Alakhras et al. (2015); Tagliafico et al. (2015a)

Retrospective studies

Five articles:

Only symptomatic women and/or women with a diagnosed breast cancer: Ariaratnam et al. (2018); Schrading et al. (2015)

Diagnostic pathway unclear: Munir et al. (2016); Waldherr et al. (2016); Freer et al. (2015)

Key findings

The available evidence only discusses one emerging technique: DBT-guided VAB. No studies reported on changes in the rate of procedural complications with DBT-guided VAB compared to other image guiding.

DBT provides clear x, y, z coordinates from first imaging and improves localisation which contributes to faster biopsy time

DBT is a helpful tool for triangulating lesions that are seen on only one view on either FFDM or DBT. This alleviates the need for additional mammographic projections for lesion localisation or confirmation. Precise targeting of a lesion's location can be calculated from one DBT projection, which is a substantial improvement over DM. Evidence from smaller retrospective studies in clinical environments suggests that DBT-guided VAB supports faster biopsy because coordinates data available more quickly and, in one study, DBT-guided VAB was up to 43% faster than DM-guided VAB.

Emerging evidence indicates that DBT-guided VAB has a superior technical success rate compared to DM-guided biopsy, resulting in reduction in avoidable harm to women from repeat biopsy

Evidence from two small retrospective studies in women with lesions suspicious for breast cancer indicate that DBT-guided VAB has a superior technical success rate compared to DM-guided VAB. Study results show that DBT-guided VAB was successful in 100% of biopsies (including those presenting with subtle mammographic findings like AD), with stereotactic-guided VAB having a slightly lower success rate.

There is moderate evidence suggesting that lesions found using DBT imaging that are occult from DM or ultrasound images should be considered suspicious for malignancy and should undergo biopsy. These findings have positive implications for workflow in terms of fewer repeat biopsies and faster procedure times.

DBT-guided VAB is emerging as a promising technique but more research is needed

DBT-guided VAB is an emerging technique. The literature comparing stereotactic-guided VAB to DBT-guided VAB is limited to a small number of little studies; however, in current literature suggests that DBT has promising potential for use in guiding biopsy (especially for lesions that are FFDM or sonographically occult). These findings are based on less than 400 biopsies, all of which investigated the use of DBT-guided VAB. In addition, this literature review found no evidence discussing avoidance of benign biopsy where calcification was the main presentation.

One study reported that women were satisfied with DBT-guided VAB

Patient satisfaction with DBT-guided VAB reported similar overall patient satisfaction as that reported for stereotactic VAB.

Systematic reviews and narrative literature reviews

Our literature search identified five narrative literature reviews (Eghtedari et al., 2018; Michell & Batohi, 2018; Destounis, 2017; Destounis et al., 2017; Mall et al., 2017) that summarised evidence on the role DBT can play in image-guided biopsies. These reviews noted that DBT-guided VAB:

- is being increasingly used as an alternative to DM-guided VAB
- is faster, has superior accuracy in lesion targeting and has a lower complication rate compared with DM-guided VAB, and
- should be carried out on findings seen only on DBT (i.e., those that are FFDM or sonographically occult) as there is a high malignancy rate amongst these lesions, and shows promise for evaluating occult lesions; however, more studies are needed to fully evaluate its efficiency and utility.

6.1. There is some evidence that DBT-guided VAB improves lesion localisation, which contributes to faster biopsy procedures

DBT-guided biopsy is a new technique and one that is essential if the suspicious lesion is FFDM or sonographically occult. When using DBT needle localisation, the reader scrolls through the 3D stack to obtain the target coordinates (x, y and z). These coordinates are available from the first DBT imaging without need to take other subsequent imaging to pinpoint the exact location of a suspected abnormality.

Systematic reviews and narrative literature reviews

No systematic reviews discussed DBT-guided biopsy.

In their narrative literature review, Mall et al. (2017) reported that the higher accuracy in lesion localisation demonstrated by DBT is important as failure to accurately locate and then obtain an appropriate tissue sample can lead to a series of negative effects including putting the patient through multiple invasive procedures and incorrect or ineffective tissue sampling.

Prospective studies

In Alakhras et al.'s prospective study of 50 cases (27 with breast cancer) (see *section 3.1* for a description of methodology), FFDM + DBT had a higher sensitivity with regards to the localisation of lesions compared to DM. Lesion localisation was indicated on a paper image of CC and MLO views and all were compared to master copies. Location sensitivity was not high with

FFDM + DBT (0.5630); however, the sensitivity with FFDM was significantly lower (0.4690, $p<.0001$), suggesting FFDM + DBT is a superior choice when it comes to determining a suspected lesion's exact location.

Tagliafico et al. (2015a) found that a small sub-population of participants ($n=3$) who underwent both procedures all reported that DBT-guided VAB was completed more quickly than DM-guided VAB but provided no further comment.

Retrospective studies

Evidence from three small (samples of less than 200 women each) retrospective studies suggests that that using DBT-guided VAB enables faster biopsy.

Both Schrading et al. (2015) and Waldherr et al. (2016) reported DBT-guided VAB procedures took significantly less time than stereotactic VAB procedures (around 15 mins per procedure compared to more than 20 minutes):

- Waldherr et al. (2016): mean procedure time of 15 mins (range, 7-28 min) for DBT-guided VAB vs. 23 mins (range, 11-46 mins) for stereotactic-guided VAB, ($p<.0001$).
- Schrading et al. (2015):
 - Average time to lesion identification: 4 mins (range: 2-12 mins) for DBT-guided VAB; 15 mins (range: 2-34 mins) for prone stereotactic-guided VAB, ($p<.001$), and
 - Mean procedure time: 13 mins (range: 8-32 mins) for DBT-guided VAB; 29 mins (range: 12-65 mins) for prone stereotactic-guided VAB, ($p<.0001$).

Schrading et al. noted that the main reason for the significant reduction in procedure time was the increased ease of target lesion reidentification, taking a mean time of four minutes (range: 2-12 mins) with DBT-guided VAB compared to 15 minutes (range: 2-34 minutes) for prone stereotactic-guided VAB, matching Waldherr et al.'s comments on precise lesion location with DBT-guided biopsy. In addition, Waldherr et al. reported the decreased time was also due to the fewer control images needed during the procedure. Overall, these are promising results indicative of potential improvements in workflow.

Munir et al. (2016) noted findings of high patient satisfaction in patients who underwent DBT-guided VAB due to the procedure being faster and less painful compared to PS VAB.

6.2. There is promising evidence that DBT-guided VAB has a superior technical success rate

Evidence from two small (samples of less than 200 women each, a limited number of procedures performed, and diagnostic pathway unknown/unclear) retrospective studies suggests that that using DBT-guided VAB enables more accurate procedures compared to DM-guided VAB.

Retrospective studies

Waldherr et al. (2016) described the differences in technique that may have resulted in the higher technical success rate for DBT-guided VAB. With DBT, readers can scroll through slices and accurately locate the lesion and precisely estimate the x, y and z (depth) co-ordinates. During stereotactic-guided VAB readers must pinpoint the same point on two geometrically different images which in situations such as diffuse lesions or distortions can result in a miscalculation. DBT-guided VAB allows accurate calculation of distance between the skin and

the lesion under compression, which enables easier planning of the safest and least invasive biopsy route.

In 2016, Waldherr et al. published a small retrospective study comparing post-implementation DBT-guided VAB to pre-implementation of stereotactic-guided VAB in a clinical setting. In this study, 148 DBT-guided VAB and 86 stereotactic-guided VABs were performed to evaluate DBT's feasibility and clinical performance. Patients who had undergone DBT-guided VAB in 2012 and 2013 were compared to data from the last complete full year in which stereotactic-guided VAB was used (2011). The authors found that all DBT-guided VABs detected and targeted AD with high accuracy and diagnosed microcalcifications with the same clinical performance as stereotactic-guided VAB. All DBT-guided VABs (100%, 124/124) and 95% (82/86) of stereotactic-guided VAB were technically successful on the first attempt. Three of the four failed stereotactic-guided VAB were successful on subsequent attempts.

DBT-guided VAB's superior performance compared to prone stereotactic-guided VAB was also demonstrated in a study by Schrading et al. (2015). Like Waldherr et al. (2016), Schrading et al. carried out a small retrospective study to compare the clinical performance of DBT-guided VAB with that of prone stereotactic-guided VAB. The authors investigated 216 suspicious findings from 205 patients who underwent image-guided biopsy. There were no differences in the cohort's age, lesion size and distribution of cancerous/benign lesions. A total of 159 women underwent prone stereotactic-guided VAB and 51 underwent DBT-guided VAB. Schrading et al. (2015) found that DBT-guided VAB either matched or outperformed standard prone stereotactic-guided VAB:

- All biopsies guided by DBT were technically successful (51 out of 51 lesions), compared to prone stereotactic-guided VAB where only 154 of 165 (93.3%) lesions were successfully biopsied
- 11 prone stereotactic-guided VAB had to be cancelled because the target lesion was not accessible or (in two of the 11 cases), prone stereotactic-guided VAB failed to sample microcalcifications correctly, and
- One of the 11 unsuccessful prone stereotactic-guided VAB was on a small cluster of low contrast microcalcifications which were successfully biopsied using DBT-guided VAB.

Schrading et al. (2015) also reported that a lower average number of procedural images was needed with DBT-guided VAB: five images (range: 4-8 images) with DBT-guided VAB compared to eight images (range: 5 to 13 images) with prone stereotactic-guided VAB. Schrading et al. noted that radiation dose increased with DBT-guided VAB; however, they did not provide MGD data (although they did note that the overall dose may be lower if fewer total images are required with DBT). While the lesion localisation and biopsy rates are promising in this study, the authors did note that the smaller number of participants that underwent DBT-guided VAB may affect the validity of their findings.

While no significance tests were performed on the technical success rates in either Schrading et al. (2015) or Waldherr et al.'s (2016) studies, the results have clinical significance. Patient tolerance and compliance for these procedures is important as an increase in pain or discomfort may increase anxiety in an already stressful setting. Failure to accurately obtain suitable tissue samples can lead to an increased number of procedures, the patient undergoing multiple invasive procedures with incorrect sampling of tissue. Therefore, the 100% technical success rate that both studies reported for DBT-guided VAB, while only slightly higher than the rate for PS VAB, may have a significant impact on both clinician and patient experience.

This literature review found no evidence discussing avoidance of benign biopsy where calcification was the main presentation.

Table 16 (below) summarises the study findings on technical success rate and procedure time.

Table 16: Primary studies comparing DBT-guided VAB to PS VAB

Study	Study details	Technical success rate	Procedure time	PPV and malignancy rate
Ariaratnam et al. (2018)	Retrospective study of 38 non-calcified DM occult lesions were biopsied and pathology results were retrospectively reviewed Unit: Hologic Selenia Dimensions	DBT-guided VAB: 38/38 successful (100%)	Mean DBT-guided VAB time: 15 min (range 10-51 min, median 13 min), 51 min outlier due to a vasovagal episode after initial positioning	DBT-guided VAB PPV: 21%
Waldherr et al. (2016)	Procedural and pathological results from 148 DBT-guided VABs and 86 PS VABs were retrospectively compared	DBT-guided VAB: 148/148 successful (100%) PS-guided VAB: 85/86 successful (99%), in 3 (4%) cases the needle to be readjusted	Mean DBT-guided VAB time: 16 min (range 7-28 min). Mean PS guided VAB time: 23 min (range 11-46 min). $p < .0001$	No data
Freer et al. (2015)	36 DM occult lesions that underwent DBT-guided VAB were retrospectively reviewed	DBT-guided VAB: 34/36 successful (94%)	Mean DBT-guided VAB time: 22 min (range 11-59 min).	DBT-guided VAB PPV: 47% (95% CI: 30.4%, 64.5%)
Schrading et al. (2015)	PS VAB was performed on 161 lesions and DBT-guided VAB on 51 lesions.	DBT-guided VAB: 51/51 successful (100%) PS-guided VAB: 154/165 successful (93%). In one of the 11 failed PS VAB, DBT-guided VAB was successfully performed	Mean DBT-guided VAB time: 13 min (range 8-32 min). Mean PS guided VAB time: 29 min (range 12-65 min). $p < .0001$. DBT-guided VAB was significantly faster at lesion targeting however, no significant difference in tissue sampling time compared to PS VAB	No data

6.3. DBT-guided VAB is effective in imaging mammographically occult lesions during biopsy

A high malignancy rate is observed in lesions that are DBT-only findings (i.e., lesions were FFDM or sonographically occult), suggesting an increased need to biopsy these findings in real-world scenarios. If no DM or ultrasound correlate exists, DBT has a role to play in the biopsy of lesions without recourse to MRI.

Retrospective studies

Both Freer et al. (2015) and Ariaratnam et al. (2018) reported good success rates using DBT-guided VAB for lesions not detected on other imaging modalities.

In 2015, Freer et al. investigated the use of DBT-guided needle localisation with surgical excision of DBT- suspicious/sonographically occult lesions. Of the 36 lesions that met the inclusion criteria, all were AD presentations (which are often subtle and difficult to detect with DM). A technical success rate of 97% (35 of 36 lesions) and a high risk of malignancy (PPV 47%, 17 of 36 lesions, 95% CI: 30.4%, 64.5%) in the abnormalities detected was found. Average lesion size was 7.9mm (range: 3-19mm). These findings show potential for radiologists to begin adopting DBT-guided procedures in cases of DM and sonographically occult lesions.

More recently, Ariaratnam et al. (2018) published a retrospective study providing further insight into DBT-guided VAB of DBT-detected, sonographically occult lesions. The authors retrospectively reviewed 1116 consecutive PS VAB performed between December 2012 and February 2016, with three readers. The cases were reviewed by two out of three independent readers separately. A total of 38 lesions were biopsied, pathologies reviewed and a PPV₃ of 26% reported. Ariaratnam et al. (2018) reported a concordance rate of 100% and a malignancy risk of 26%. While the risk of malignancy was lower than that found by Freer et al. (2015), it falls within published guidelines (26-32%) and if all actionable pathologies were included it increased to 58%. The high risk of malignancy in these studies for lesions only detected by DBT confirmed that routine biopsy is needed for these specific cases.

6.4. One study reported that women were as satisfied with DBT-guided VAB compared to DM-guided VAB

One prospective study reported on patient satisfaction with DBT-guided VAB.

Tagliafico et al. (2015a) investigated the short-term quality of life (QoL) effects of image-guided VAB comparing DBT and DM. Ninety participants who underwent VAB at a medical centre (diagnostic pathway unknown) between November 2014 to March 2015 were asked to complete a survey. Surveys were collected from 45 women who had undergone DBT-guided VAB and 45 who had undergone DM-guided VAB. The survey used a slightly modified questionnaire of a previously validated survey assessing short term QoL aspects related diagnostic procedures¹⁶ (see box, right). The survey assessed 10 attributes, six of which were directly related to the procedure. Tagliafico et al. found that patients undergoing DBT-guided VAB had a decreased short-term QoL and the procedure was less well-tolerated compared to DM-guided VAB; however, the overall satisfaction was similar between the two procedures with a slight favouring of DBT-guided VAB compared to DM (95.2% vs 90.1%, $p < .02$). Additionally, there was no significant difference between the six individual survey items relating to the procedures, suggesting that no one area of the DBT-guided VAB procedure was specifically worse than the DM-guided VAB: patient unhappiness was due to the overall experience of undergoing biopsy. The authors reflected that, due to gaps in the literature at the time of publication, they were not aware of which changes in an outcome a patient would identify as being important. Therefore, while these are statistically significant findings, they may not be significant in a clinical sense.

¹⁶ Survey attributes assessed: Pain or discomfort before the test, Pain or discomfort during the test, Fear or anxiety before the test, Fear or anxiety during the test, Physical function after testing, Mental function after testing.

7. IMPLEMENTATION INTO THE BREAST CANCER ASSESSMENT CENTRE

There currently is no international or Australian guidance about which patients should be imaged with DBT during recall for assessment or diagnosis or which DBT imaging protocol (alone or with FFDM, in one or two views, with/without s2DM) is preferred. In Australia, some BSA assessment centres provide DBT as part of the imaging work-up for all non-calcified lesions, other centres rely on the attending radiologist to determine imaging requirements and some centres do not have DBT capability. DBT has been implemented into clinical practice in other single institution clinics and national screening programs in the United Kingdom and Ireland. Evidence about key implementation considerations is an important component of understanding the evidence about DBT's role in the breast cancer assessment centre.

Key findings from Allen + Clarke's literature review on the role of DBT in screening

Several issues need to be considered to ensure the maximum benefits of DBT are realised before it is implemented as a preferred screening strategy.

- DBT requires additional time per view to acquire images. This ranges from an additional 10 seconds to up to one additional minute.
- Interpretation times have been shown to increase with DBT as readers need to scroll through up to 100 DBT images compared to four FFDM images, but the interpretation times tend to decrease as readers gain experience in interpreting DBT. Reading times overall remain higher compared to FFDM. While the scrolling approach may increase reading speed, studies reported that reading a DBT stack takes at least twice as long (up to 90 seconds).

Detailed analysis of the cost effectiveness has not been performed in jurisdictions other than the United States. American modelled analyses show that DBT is cost-effective (in terms of finance) for community-based practices (although these are very specific to the insurance programs).

Overall results on PPV₂ and PPV₃ indicated that FFDM + DBT was more accurate than FFDM when used as a basis for recommending or performing biopsies. PPV results for DBT + s2DM are also promising but present more varied effect size than results for FFDM + DBT; however, the direction of effect indicates that DBT + s2DM might be useful in terms of reducing false positives leading to either recall or biopsies.

The PROSPECTS trial and the Maroondah trial will further investigate the cost effectiveness of using DBT in breast cancer screening (as a primary screening tool) and in the assessment of suspicious lesions. The PROSPECTS trial will be conducted over seven years from 2018, with initial results to be presented within 18-24 months. Maroondah will report earlier than this date. Both trials will provide useful information about implementation considerations which will be useful advice for national and state screening programs.

Studies reported on changes made to clinical practice after DBT had been implemented. Many of these studies are set in the United States, where screening mammography is usually completed annually, and images are read by one radiologist. This is a different context to how the BSA program is implemented. Regardless, all studies noticed positive changes in workflow in terms of reduced diagnostic imaging and more efficient clinical pathways to biopsy due to increased PPV₃ at screening or follow-up imaging. There is also a growing body of evidence that seeks to optimise the use of DBT (such as studying different combinations of DM and DBT to minimise radiation dose while still acquiring clear images) and research on optimising workflow in

screening and assessment clinics. This work is ongoing and no definitive way to use DBT has emerged yet.

7.1. No research relevant to the Australian context was identified on the incremental costs associated with implementation

No studies reporting on the cost of implementing DBT into a breast care clinic were identified.

7.2. Workflow could be improved with the implementation of DBT, but this depends on several practices

This literature review describes findings from five literature reviews and six primary studies which described implementation issues associated with DBT in the assessment and diagnostic environment. Primary studies already incorporated into systematic or literature reviews were reviewed but not separately assessed unless additional material not described in the systematic review or narrative literature review was included in the primary study. Most of these studies included mixed study populations (i.e., women recalled to assessment and symptomatic women or symptomatic women only or insufficient information about the diagnostic pathway).

The studies discussed in this literature review are listed below.

Systematic and/or literature reviews

Systematic reviews: none

Five narrative literature reviews: Destounis et al. (2017); Odle (2016); Peppard et al. (2015); Conant et al. (2014); Lee & Lehman (2013)

RCTs and prospective studies

RCTs: none

Prospective studies: none

Retrospective studies

Eight retrospective studies: Dibble et al. (2018); Balleyguier et al. (2017); Ni Mhuircheartaigh et al. (2017); Raghu et al. (2016); Bansal & Young (2015); Freer et al. (2015); Brandt et al. (2013); Hakim et al. (2010)

Key findings

Assessment centre workflow may be positively impacted by the implementation of DBT, but this depends on several practices

Implementing DBT affects workflow but the overall impact is driven by when in the screening and assessment process DBT is implemented.

If used as a screening tool, DBT may result in reduced recall to assessment imaging (or reduced further diagnostic work-up) due to improved lesion conspicuity and more accurate initial reading of screening mammograms. This could result in fewer inconclusive screening results and decrease the necessity of short-term follow-up studies; however, this is likely to be of limited benefit in the BSA program as short-term follow-up is not a preferred management technique (i.e., the program aims to resolve all results to normal/benign or malignant). DBT when used in screening may result in fewer false positive recalls, resulting in reduced work-up

of benign final outcome lesions, which could offset the lengthier image interpretation times associated with DBT; however, this literature review did not explore the full pathway of possibilities given its focus on the role of DBT in the breast cancer assessment centre.

If DBT can acquire sufficient diagnostic data in a single compression (if using FFDM + DBT), it potentially results in:

- faster diagnostic work-up because women do not have to have further mammographic work-up (i.e., reduced positioning time compared to additional DM views, fewer positioning errors and reduced need for repeat imaging, and reduced time needed to move between different types of imaging equipment)
- shorter overall additional time under compression for women, and
- fewer total diagnostic imaging studies or lesion localisation views.

DBT's impact on the need for other work-up views (including ultrasound or MRI) remains unclear, with some early studies indicating a small decrease in the use of ultrasound following implementation of DBT into the assessment centre, but others not reporting a significant decrease (i.e., ultrasound is still required, especially for the work-up of masses).

Finally, DBT has a higher PPV₃ resulted in more targeted biopsy and, when used with VAB, faster biopsy times.

Reading times for DBT are longer compared to DM images but this time could be offset by a reduction in the number of women being recalled to assessment

DBT requires larger data storage compared to DM modalities. Images take much longer to read compared to DM simply because there are many more images to scroll through; however, overall, clinics might find that the extra time is not much of an issue if fewer women are recalled to assessment (potentially allowing screening resources to be reassigned to assessment settings).

7.2.1. Improved lesion conspicuity with DBT results in fewer women requiring additional work-up views (including possible changes to the use of ultrasound)

Narrative literature reviews

While outside of the exact scope of this literature review (DBT's role in assessment), the four narrative literature reviews commented on the workflow implications if DBT is used as a primary screening. That is, DBT:

- improves lesion conspicuity which
- leads to more accurate reader decision-making and fewer inconclusive mammogram results or recommended short-term follow-up imaging studies and increased avoidance which
- results in reduced assessment/diagnostic work-up, especially of unnecessary work-ups for final benign outcomes (see *section 3.4*), and
- reduced overall reading volume as fewer false positive images are recalled to assessment which can offset the lengthier image interpretation times.

Specifically, Destounis et al. (2017)¹⁷ commented on a retrospective assessment study (citing Philpotts et al., 2013), which reviewed the initial impact DBT had on examination success when implemented into a clinical setting. While it is not clear from Destounis et al.'s paper (or the conference abstract for this study) whether DBT was used alone or as FFDM + DBT, the authors reported that after the implementation of DBT, fewer women attending for screening required additional imaging (a decrease of 72%).

A study by Chudgar et al. (cited in Michell & Batohi, 2018) reported that DBT increased diagnostic accuracy and had resulted in a decrease in the impact of MRI data on the management of multifocal breast cancers. No further information was provided about this study.

Retrospective studies

Two early studies (Brandt et al., 2013; Hakim et al., 2010) discussed the impact of DBT on the need for ultrasound imaging during assessment.

In a group of 146 women recalled to assessment for non-calcified findings suspicious on screening mammography, Brandt et al. (2013) assessed if DBT could be used instead of DM (see *section 3.1.2* for a discussion of this study's methodology). This study did not directly compare DM + DBT to DM + ultrasound, but it is included here because it described DBT's performance when ultrasound was included in the work-up pathway (reflecting real-world clinical practice). Readers read abnormal screening mammograms, comparison mammograms and DBT images. Ultrasound images were also available if required. In this study, when using DBT images, readers still required the use of ultrasound images:

- In clinical work-up, ultrasound was requested in 49% of cases.
- When DBT was used, ultrasound was required in slightly fewer cases (33-55% of cases, depending on the reader) but it is still needed.

Brandt's et al.'s research indicated that DBT might provide a more complete imaging work-up and may result in a reduction in the use of additional ultrasound views (which has important implications for assessment centre workflow); however, there is obviously still a strong role for ultrasound.

Hakim et al. (2010) asked radiologists to report on whether ultrasound imaging could be eliminated on the grounds that the DBT images provided adequate diagnostic information. The study dataset included women recalled to assessment and symptomatic women. In 32% of cases, at least one of the four readers considered that DBT provided adequate diagnostic information and that further imaging with ultrasound was not required. These lesions were benign masses with margins that were clearly seen on DBT; however, further ultrasound imaging was still needed in most cases to confirm diagnosis. This finding is consistent with Brandt et al. (2013). Conversely, this could see a significant decrease in the need for ultrasound; however, on their own, these findings are not sufficient to recommend practice change.

¹⁷ Screening and diagnostic FFDM examinations were reviewed for the one year before (August 2010 to July 2011) and one year after (March 2012 to February 2013) the implementation of DBT. A total of 9462 screening and 4611 diagnostic exams occurred during the pre-DBT period and a total of 11,101 screening and 5357 diagnostic exams occurred in the post-DBT period. Destounis et al. (2017) also described changes in workflow from a screening perspective but these are not described here.

7.2.2. DBT acquires sufficient data in a single compression leading to faster work-up of suspicious lesions

Narrative literature reviews

Four narrative literature reviews identified in our literature search provided commentary on workflow. Odle (2016) and Lee & Lehman (2013) summarised the implementation and workflow implications of DBT as follows:

- With DBT, sufficient diagnostic data can be acquired in a single compression (if using FFDM + DBT) resulting in faster diagnostic work-up, particularly:
 - shorter overall additional time under compression for women
 - fewer positioning errors compared to DSCV
 - less patient and health practitioner time spent completing or participating in more than one imaging protocol (eg, FFDM + ultrasound or MRI examinations)
 - fewer diagnostic imaging studies (eg, additional DSCV or lesion localisation views), and
 - Faster biopsy times (as described in *Chapter 5*)

Destounis et al. (2017) reported decreases in the number of diagnostic views required to complete the diagnostic work-up per patient (a decrease of 11% overall and a decrease of 57% for DSCV).

Peppard et al. (2015) reported on individual clinic experience using DBT in a diagnostic setting. Peppard et al.'s clinic initially used both DBT and FFDM to evaluate non-calcified and one view findings (i.e., when a suspected abnormality was only visualised on DM_{MLO} or DM_{CC}). In their clinical experience, it became apparent that DBT was adequate for the evaluation of these findings (i.e., DSCV/DM images did not provide additional diagnostic information to that available with DBT). Subsequently, the number of supplemental DM images required decreased due to the improved lesion localisation that DBT provided. Conant et al. (2014) reported:

"Incorporating DBT in the diagnostic, or problem-solving imaging of patients has the potential to limit, or possibly replace, much of the additional views performed decreasing the x-ray dose and time of imaging".

Further information on DBT's performance in comparison to DSCV is provided in *section 3.1*.

Retrospective studies

As noted in *Chapter 3*, several institutions have replaced DSCV with DBT, with associated gains in workflow efficiency.

After introducing DBT into the workup of screen-detected abnormalities at their screening centre, Ni Mhuirheartaigh et al. (2017) commented on their confidence in replacing DSCV in the assessment of non-calcified masses, asymmetries, and distortions with DBT without any loss in diagnostic performance.

Bansal & Young (2015) also commented how their radiographers replaced DSCV with FFDM + DBT for assessing non-calcified soft tissue abnormalities, as they found DBT to be easier to use and does not require the precise positioning needed to get accurate information from DSCV. Dibble et al (2018) reported a reduction in the need for further diagnostic imaging (i.e., women recalled to assessment may only need additional ultrasound imaging rather than DM + ultrasound).

Two studies commented on PPV₃, with both reporting increases in PPV₃ with the use of DBT (Raghu et al., 2016; Freer et al., 2015). Increases in PPV₃ are important because it indicates that biopsy is better targeted towards malignancy and therefore service efficiency is increased (with commensurate increases in workflow and reduction in avoidable patient harm).

Together, the limited available evidence base indicates that there could be significant changes in the clinical pathway for the assessment and diagnosis of breast cancer if DBT is adopted as a screening strategy. That is:

- fewer women are likely to be recalled to assessment because of more definitive screening imaging
- those who are recalled are more likely to require less DM and (possibly) ultrasound imaging (although ultrasound will still be important) if the DBT images contain sufficient diagnostic and lesion localisation information to determine either that biopsy is needed (and coordinates are already available), and
- biopsy may be better targeted resulting in avoided unnecessary biopsy or avoided repeat biopsy (and less patient harm).

However, as demonstrated throughout this literature review, there is still no clear evidence about the optimal way to implement and use DBT in the diagnostic and assessment.

7.2.3. DBT images take longer to read than DM images

Allen + Clarke's literature review on the role of DBT in screening found that, due to the increased number of images readers must look at, a set of DBT images (using a two-view bilateral approach) takes longer to interpret than a set of DM images for the same imaging sequence. There appears to be no way around this increase in interpretation time, even as readers become more proficient at reading these types of images because of the simple fact that DBT creates more images for readers to assess. For example, for a single breast compressed to 5cm thick which has undergone bilateral FFDM + DBT there may be up to 100 images for a reader to review. Therefore, the question considered in some research is in what way should DBT be used to maximise both accuracy and time efficiency.

Tagliafico et al. (2017) discussed the time requirements for each of six reading strategies. While the double reading of the DBT + s2DM combination was found to have a significantly greater reader performance than any of the other five combinations, on average it took almost 60% longer than the next slowest combination, and five times longer than the fastest imaging strategy:

- s2DM (single read): 31 – 34 seconds
- DBT (single read): 43 – 47 seconds
- DBT + s2DM DBT (single read): 75 – 79 seconds
- s2DM (double read): 65 seconds
- DBT (double read): 90 seconds, and
- DBT + s2DM (double read): 154 seconds.

This increase in reading time is consistent with studies discussed in *Allen + Clarke's* literature review on DBT in screening.

Balleyguier et al. (2017) evaluated whether the concurrent use of CAde could positively impact on reader performance and reader time. In this multi-centre study, six readers (all radiologists)

reviewed 80 bilateral screening and diagnostic work-up cases ($DM_{CC} + DBT_{MLO}$) using GE SenoClaire with and without CAdE. A total of 45 cases had breast cancers ($n=23$) or benign lesions. Mixed results were reported with CAdE: some readers improved their performance and two did not. There was limited difference in AUC measurement for $DM_{CC} + DBT_{MLO} + CAdE$ compared to $DM_{CC} + DBT_{MLO}$: the average AUC measurement was 0.850 compared to 0.854 respectively (i.e., CAdE was not inferior to $DM_{CC} + DBT_{MLO}$). If there was no increase in diagnostic accuracy, it is important to consider impact on reading time. Balleyguier et al. reported that CAdE resulted in faster average reading: 48.2 sec without CAdE compared to 39.1 sec with CAdE. The average reduction in reading time was 23.5%. Overall, the CAdE-enhanced images increased reading performance without a drop in diagnostic accuracy.

While DBT increases interpretation time for readers, it also has better accuracy compared to FFDM. Therefore, these factors need to be weighed against each other to decide which is most important in a clinical assessment setting. Previous studies and narrative reviews have shown that while there may be an increased burden on readers and a learning curve for clinics to overcome, clinics can end up more efficient in the assessment area as fewer follow-up images (and therefore less overall reading) is required post-implementation of DBT, which also results a lower total radiation dose for patients and reduced biopsy imaging.

7.3. Reader performance

Findings from *Allen + Clarke's* literature review on the role of DBT in screening

Most studies involve readers (almost always radiologists) who have a range of experience in breast screening and radiology in general. Interpretation accuracy improved more for less experienced readers compared to those with more experience. It is unclear whether this improvement reflects the development of less experienced practitioners' competence in breast imaging in general, or whether the increased lesion conspicuity seen with DBT makes it 'easier' to read and detect subtle findings without as much overall clinical imaging experience.

Evidence suggests that diagnostic accuracy (as measured by area under the ROC curve, sensitivity, specificity and PPV/NPV) increases with the use of DBT or is at least equivalent to FFDM (as described in *section 3.1*). While this literature review returned many studies looking at overall diagnostic accuracy, only two narrative literature reviews and seven studies specifically provided stratified analysis of the relationship between reader experience prior use of FFDM/DM or DBT and diagnostic performance.

Systematic reviews and narrative literature review

Systematic review: none

Two narrative literature reviews: Michell & Batohi (2018); Peppard et al. (2015)

RCTs and prospective studies

RCTs: none

Prospective studies (diagnostic pathway unknown): Alakhras et al. (2014)

Retrospective studies

Six retrospective studies:

Women recalled to assessment only: Heywang-Köbrunner et al. (2017); Tucker et al. (2017)

Women recalled to assessment and symptomatic women: Wallis et al. (2012)

Symptomatic women only: Mariscotti et al. (2016); Seo et al. (2016)

Diagnostic pathway unknown: Thomassin-Naggara et al. (2015)

Key findings

All readers improve their diagnostic accuracy when using DBT but there is some evidence to suggest that less experienced readers improve more DBT compared to more experienced readers

Consistent evidence (mostly based on small numbers of cases and readers but including a large sub-study from the TOMMY trial) indicates that all readers improve their performance with DBT, but less experienced readers improve more when using either FFDM + DBT or DBT_{ML0/CC} compared to FFDM or DSCV. Three retrospective studies reported that greater gains tend to accrue to less experienced readers. While different definitions are used to describe “more” or “less” experience (including prior DBT experience and prior DM experience by volume or time), the trend is the same. Possible reasons for the increase in less experienced readers may be due to improved lesion conspicuity with DBT, making it easier to detect abnormalities presenting as subtle mammographic findings (especially AD and asymmetry) which may be missed on FFDM by less experienced readers. Only one retrospective study reported mixed results with some readers performing better with FFDM compared to FFDM + DBT; however, several important bias within the study were not accounted for.

More research is needed to validate these findings

More studies are needed to validate the influence of prior DBT/DM experience and confirm that it is a combination of prior experience with DM that drives diagnostic accuracy gains (i.e., readers transition to DBT smoothly with only a small amount of additional training) and the influence of improved lesion conspicuity for those with less experience who can then better visualise subtle findings, improving detection.

7.3.1. All readers improve their diagnostic accuracy when using DBT but there is some evidence to suggest that less experienced readers improve more DBT compared to more experienced readers

Systematic reviews and narrative literature reviews

None of the systematic reviews presented diagnostic accuracy results stratified by experience.

Michell & Batohi (2018) briefly reported that there are differences in reader performance with the use of DBT compared to other mammographic imaging. Citing studies by Tucker et al. (2017) and Wallis et al. (2012), Michell & Batohi noted that there is a trend for specificity and sensitivity to increase, with greater performance improvements (as measured by sensitivity and specificity) seen in readers with less experience. Tucker et al. and Wallis et al. are described in further detail below.

Prospective studies: diagnostic pathway unknown

The literature search identified one prospective study that stratified diagnostic performance with DBT by prior experience of using DBT. Alakhras et al. (2014) completed a small study (27 breast cancer cases, 23 benign/normal cases chosen because of indeterminate mammographic findings) comparing reader performance with FFDM + DBT to FFDM as a function of radiologist experience. The 26 readers were grouped according to prior experience with DBT: nine readers had no experience, nine readers had up to 12 hours workshop training on DBT, and eight readers had read up to 2000 DBT cases per annum. All readers read at least 1000 mammograms per annum, with 24 readers reading more than 2000 mammograms per annum. Diagnostic performance (measured by sensitivity, specificity, AUC and JAFROC FOM) improved for all groups of readers when using FFDM + DBT:

- Readers with no DBT experience: AUC measurement (FFDM compared to FFDM + DBT) increased from 0.682 to 0.775, ($p=.004$)
- Readers with up to 12 hours training: AUC measurement (FFDM compared to FFDM + DBT) increased from 0.680 to 0.790, ($p=.004$)
- Readers with clinical experience using DBT: AUC measurement (FFDM compared to FFDM + DBT) increased from 0.681 to 0.789, ($p=.042$).

There is no difference between groups based on prior DBT experience (although all readers were experienced in breast imaging). Alakhras et al.'s concluded that diagnostic performance for all readers increased regardless of prior DBT experience, suggesting that readers can quickly adapt to using DBT and achieve gains in diagnostic accuracy.

Retrospective studies of women recalled to assessment

Improved accuracy for all readers with the greatest gains favouring readers with less than 10 years' experience is established in a retrospective sub-study of TOMMY trial data (5888 FFDM and 4608 FFDM + DBT cases chosen to represent a wide range of mammographic presentations). Tucker et al. (2017) compared the diagnostic performance of 24 readers (comparing FFDM to FFDM + DBT imaging with readers blinded to clinical outcome). Readers were very experienced: the median years' experience was 10 years and the median number of mammograms read per annum was 8000 cases. When using FFDM + DBT:

- sensitivity improved for 21 readers but decreased slightly for three of the most experienced radiologists
- specificity significantly improved for all readers regardless of experience ($p<.0001$)
- there was a larger statistically significant gain in sensitivity and specificity for readers with less than 10 years' experience (increasing from 86% to 91% and 55% to 69% respectively, $p=.03$) compared to more experienced readers (equivalent sensitivity at 88%; specificity increases from 53% to 65%), and
- there was equal sensitivity (88%) with the two imaging protocols but higher specificity for readers with more than 10 years' experience (53% with FFDM compared to 65% with FFDM + DBT).

Tucker et al. concluded that gain in sensitivity was possibly due to improved lesion conspicuity of subtle mammographic presentations like AD or asymmetry (as discussed in *Part 4*). The authors also noted that increases in sensitivity and specificity with FFDM + DBT may be underestimated in this study because all participants were cases recalled to assessment (and therefore had an abnormality visible on FFDM, which readers were aware of).

Other studies of women recalled to assessment assessed reader performance using DBT_{MLO} (either in a wider imaging protocol, adjunctive to FFDM or alone). Heywang-Köbrunner et al., (2017) reported on 16 radiologists experienced in screen reading (although the scope of experience was not fully described); two readers had some DBT experience (>500 cases read in a clinical setting), but the other readers did not. Images included masses, AD, global/focal asymmetry (i.e., reflective of a range of mammographic findings expected in a clinical environment). DBT images were read with FFDM + ultrasound and compared to FFDM + DSCV + ultrasound. Using AUC measurement as the comparative measure:

- 10 readers performed better with the FFDM protocol (screening mammogram, ultrasound and additional mammographic views) (delta range: -0.011 to -0.110)
- One reader performed as well with both protocols (delta range: 0)
- Five readers performed better with the DBT protocol (screening mammogram, ultrasound and DBT images) (delta range: 0.027 to 0.101).

This study demonstrates that there was a degree of difference in reader performance which may reflect low experience with DBT before participating in the study as the two readers with DBT experience reported higher AUC with the DBT protocol. In a sub-analysis, Heywang-Köbrunner reported that seven 'regular assessors' (not further defined) reported slightly higher (but not significant) AUC with the DBT (0.011 compared to -0.024, $p=.09$). These results are consistent with Dibble et al.'s findings that more experienced readers had higher increased but are not consistent with the other studies reported in this literature review. The authors explained that there was a potential bias influence: four readers read a double set of images and had lower AUC scores with the DBT protocol. Further information about experience was not provided, which makes it difficult to determine the full impact of this bias. Given these issues, Heywang-Köbrunner et al.'s results should be treated with some caution.

Retrospective studies of women recalled to assessment and symptomatic women

One study reported on DBT alone using a photon-counting DBT unit (Wallis et al., 2012). Images came from symptomatic women or those recalled to assessment. 10 radiologists had more than 10 years' experience; 10 radiologists had less than 10 years' experience. Comparing FFDM to DBT, ROC analysis showed that less experienced readers had statistically significantly lower FOM values for FFDM compared to more experienced readers (AUC measurement 0.760 compared to 0.787) but that they improved more when using DBT (increasing to 0.779 compared to 0.770 for more experienced readers). That is, use of DBT only increased diagnostic accuracy for less experienced readers (although these readers are quite experienced). It is unclear how much impact the use of a DBT prototype unit compared to the commercial units used in later studies.

Retrospective studies of symptomatic women

Seo et al. (2016) reported JAFROC FOM values for each of six radiologists, with experience ranging from 2 to 15 years (not further described). Images were drawn from a consecutive cohort of 219 women presenting at a single institution with symptoms or suspicious mammographic findings (all types). While the authors did not analyse findings, table data suggested that, for all readers, diagnostic accuracy and sensitivity increased with the use of FFDM + DBT compared to FFDM (with significant increases for all but the radiologist with 15 years' experience); however, there was no clear difference in magnitude of improvement by years of experience in breast imaging. Mixed, non-significant changes were reported in specificity for individual readers.

Mariscotti et al. (2016) compared inter-reader agreement for the detection of ILC (a cancer that can be hard to detect on FFDM due to its subtle mammographic presentation) on FFDM + DBT compared to FFDM. The 12 readers were experienced, having read at least 3000 mammograms and at least 1000 DBT cases per annum. Inter-reader agreement between readers with less than two years' experience ($n=4$) and those with 3-16 years' experience ($n=8$) showed improvement in detecting ILC with FFDM + DBT compared to FFDM. Incremental improvement in AUC measurement was 0.10 ($p<.0001$) for less experienced readers compared to 0.04 ($p<.002$) for those with more than two years' experience. While diagnostic accuracy for both groups increased, the less experienced readers experienced greater gains in detecting ILC with FFDM + DBT. This study shows that DBT may dampen the effect of reader inexperience, which is particularly important if cancers have a subtle presentation.

Retrospective studies: diagnostic pathway unknown

Thomassin-Naggara et al. (2015) retrospectively evaluated four readers' performance in assessing 155 FFDM and DBT_{MLO} images (61 breast cancers; 94 benign or high-risk lesions). Two radiologists were experienced (10 or 30 years' experience); two had less than three years' experience. Readers had access to clinical breast examination data and prior DM images. For lesion characterisation, there was a small difference in kappa value for senior/junior readers with FFDM images but almost no kappa difference between the four readers when DBT images were reviewed. Thomassin-Naggara et al. reported that FFDM + DBT improved the accuracy for all readers (in terms of sensitivity and NPV), but the greatest improvements in diagnostic accuracy were achieved by less experienced readers. Statistically significant increases in AUC measurement for the two less experienced readers were 0.155 and 0.157 with FFDM + DBT compared to increases of 0.108 for one of the more experienced readers; the other experienced reader reported a non-significant increase of 0.051.

7.3.2. As readers become more familiar with reading DBT images, they may come to prefer it to DM

Narrative literature reviews

Peppard et al. (2015) summarised early literature on reader preference for adjunctive DBT in the evaluation of non-calcified lesions. Citing Brandt et al. (2013), Zuley et al. (2013), Tagliafico et al. (2012), Hakim et al. (2010) and Teertstra et al. (2010), Peppard et al. noted that readers preferred to use DBT compared to other mammographic work-up views. In these studies, as readers became more familiar with DBT, their preference for it increased. While not directly linked to performance, Peppard et al. implied that performance improved with reader experience and growing preference for DBT.

7.4. Reader training in DBT interpretation is required for reading screening and assessment settings

As more screening clinics and assessment centres report experience with DBT in the real world, they are reporting that, as with all new imaging technologies, there is a learning curve which usually causes an initial disruption to clinic processes and scheduling. However, most clinics find they overcome this relatively quickly and end up more efficient than prior to implementation. Additionally, there are currently no set guidelines on the optimal method to use DBT to get the best performance out of readers.

Currently, to become qualified to read DBT images in a clinical setting there is a requirement to undergo eight hours of training. Conant et al. (2014) felt this level of training may not be

adequate as in their practice they noticed in the initial period after implementation an increase in the DBT-only cancers. However, some readers began recalling patients with subtle distortions that were only seen on the DBT image, thinking that they had detected a DBT-only cancer. These findings were small benign masses such as cysts or newly unmasked intramammary lymph nodes. In addition, Rafferty et al. (2013) completed a second study due to a reader missing an ILC on DBT which was detected on FFDM, suggesting that initial training may not have been sufficient. Conant et al. recommended that readers need to reset their thresholds for calling back these types of cancers to prevent an increase of false positive call backs. While these findings relate to a screening setting, they also have relevance to assessment centre practice too if these types of suspicious lesions are then worked-up with further ultrasound or biopsy.

Further, Conant et al. (2014) explained how they implemented DBT initially into screening as a method of increasing reader recognition of what “normal” looked like on DBT, before using it in an assessment context. Using this staged approach, the authors estimated it took approximately 1000 readings before readers were confident in their ability to read and accurately assess the images.

Further information on reader performance is provided in *section 3.2* and *section 3.3*.

7.5. DBT images are larger and IT requirements for storage and transmission need to be considered

An issue raised in the 2016 NHS position statement on tomosynthesis was the increased file size associated with DBT images compared to DM. These images are stored in a standard DICOM format known as BTO or in a CT format. File sizes are approximate 1GB but can range from 300MB to 2.4 GB (depending on breast thickness), which may have implications for local and archival PACS storage and transmission of images to other health providers. Therefore, before any site begins clinical implementation, preparations must be made to accommodate and transmit the large file sizes. Lossless compression programs may be used to reduce the total file size without losing image quality. Some compression algorithms can achieve 4:1 lossless compression ratio which is still on the high end of general DM file size ranging from approximately 35 to 58 MB per view.

8. DIAGNOSTIC PERFORMANCE OF OTHER ADJUNCTIVE IMAGING IN THE ASSESSMENT OF LESIONS SUSPICIOUS FOR BREAST CANCER

This section discusses the diagnostic performance of:

- adjunctive DBT (FFDM + DBT) compared to adjunctive ultrasound (FFDM + ultrasound) in the assessment of lesions suspicious for breast cancer, and
- adjunctive DBT (FFDM + DBT) compared to adjunctive ultrasound (FFDM + ultrasound) in the assessment of lesions suspicious for breast cancer, and

8.1. Adjunctive ultrasound

While there is evidence that ultrasound could be used as a supplemental cancer detection test for women who have dense breasts and another risk factor for breast cancer, there is limited evidence to suggest that ultrasound be used as a primary screening tool for the early detection of breast cancer in women who have an average-risk of developing breast cancer. This is due to the high number of false positives and the lack of evidence that supplemental testing improves breast cancer mortality. Interim data from the ASTOUND trial comparing the performance of DBT and adjunctive ultrasound (i.e., DBT + ultrasound as the screening test) was reviewed but not included in this literature review because study results reported to date focus on ultrasound's role as a primary screening test. Further results relevant to assessment or diagnosis may be published with final results.

Ultrasound is not used as a supplemental screening test in the BSA program, but it is a highly useful imaging technique if a woman is recalled to assessment and it is used at all BSA centres. Ultrasound is an important part of imaging work-up to confirm or dismiss breast cancer (particularly suspicious lesions presenting as masses). Ultrasound provides further information to establish the nature of a mass or a soft tissue lesion (which may then need further mammography work-up). It is used to guide interventional procedures (eg, tissue sampling) or to further assess additional lesions identified on MRI (second-look ultrasound). Acquiring accurate images using breast ultrasound can be time-consuming and is dependent on sonographer/operator skill.

This literature review describes findings from five studies that investigated the diagnostic accuracy of DBT (FFDM + DM) compared to ultrasound (DM + ultrasound). Detailed study results are included in *Table 17 (A-C)*.

Systematic reviews and meta-analysis

None.

RCTs and prospective studies

RCTs: none

Two prospective studies:

Women recalled to assessment and women with symptoms: Kim et al. (2017)

Women presenting for screening and women with symptoms: Thibault et al. (2013)

Retrospective studies

Three retrospective studies:

Only women recalled to assessment: Lee et al. (2016)

Women presenting for screening and women with symptoms: Kim et al. (2015)

Diagnostic pathway unclear: Elizalde et al. (2016)

Key findings

This literature review identified limited evidence discussing the role of adjunctive DBT (i.e., FFDM + DBT) compared to adjunctive ultrasound (FFDM + ultrasound) in the work-up of screen-detected breast cancers. Identified literature included populations undergoing a range of screening and/or diagnostic pathways and a range of different imaging view combinations.

DM + DBT does not have inferior diagnostic accuracy compared to DM + ultrasound, but there is limited evidence about DBT's role in a work-up pathway compared to ultrasound and there is no suggestion that DBT should replace ultrasound views during work-up

Two prospective studies reported on different imaging protocols and study populations (either comparing DM + DBT to DM + ultrasound or comparing a range of techniques) making it challenging to compare findings. Both prospective studies were powered to determine the non-inferiority of adjunctive DBT compared to adjunctive ultrasound and, despite some mixed findings on AUC measurement, non-inferiority was confirmed. Retrospective observational studies reported similar results to the prospective studies: adjunctive DBT was not inferior to adjunctive ultrasound but the superiority of adjunctive DBT has yet to be demonstrated. This evidence does not suggest that FFDM + DBT should replace FFDM + ultrasound in the work-up of suspicious breast cancers. It is not possible to assess whether there are some situations in which FFDM + DBT would be preferred over FFDM + ultrasound.

8.1.1. DM + DBT does not have inferior diagnostic accuracy compared to DM + ultrasound, but there is limited evidence about DBT's role in a work-up pathway compared to ultrasound and there is no suggestion that DBT should replace ultrasound views during work-up

Systematic reviews and narrative literature reviews

None of the systematic reviews or narrative literature reviews compared the diagnostic performance of FFDM + DBT to DM + ultrasound in the assessment of breast cancer. While Lei et al.'s (2014) systematic review discussed findings from Thibault et al.'s (2013) prospective multi-site analysis of women at average-risk of developing breast cancer (discussed below), it did not explore comparisons of performance for FFDM + DBT and FFDM + ultrasound.

Prospective studies in mixed populations

Two prospective studies reported on diagnostic accuracy for adjunctive DBT compared to adjunctive ultrasound. Both studies used DBT imaging in different combinations in slightly different populations of women:

- In Kim et al.'s (2017) non-inferiority reader study, women presenting for screening and symptomatic women had FFDM + DBT (images acquired in combo mode) + ultrasound with study analysis based on FFDM + DBT images compared to FFDM + ultrasound images.

- Thibault et al.'s (2013) study of women recalled to assessment and symptomatic women had five reading arms:
 - FFDM/DSCV
 - FFDM/DSCV + ultrasound
 - FFDM/DSCV + ultrasound + DBT_{MLO}
 - DBT_{MLO} alone, and
 - DBT_{MLO} + DM_{CC}.

This section reports on Thibault et al.'s results on the differences between FFDM/DSCV + ultrasound compared DBT_{MLO} + DM_{MLO} as these are the most relevant study reading arms for this literature review research questions. Results from all study arms are discussed in *Table 17*.

In Thibault et al.'s study (which used a prototype GE Senographe DBT unit), diagnostic accuracy favoured DBT_{MLO} + DM_{CC} over DM + ultrasound (the AUC measurements were 0.7795 compared to 0.7697 respectively). Interestingly, the one view imaging protocol (DBT_{MLO} + DM_{CC}) had a higher overall AUC measurement compared to a 'fuller' protocol of FFDM + DBT_{MLO} + ultrasound (AUC measurement = 0.7628) but there was little difference in any of the imaging protocols overall. Results for sensitivity, specificity, NPV and PPV were consistent with Kim et al.'s results: lower sensitivity and NPV with adjunctive DBT but higher specificity and PPV, meaning more false recalls with the ultrasound combination. Thibault et al. concluded that FFDM/DSCV + ultrasound was not superior in terms of diagnostic accuracy to DBT_{MLO} + DM + ultrasound using bootstrap analysis of the non-parametric comparisons of AUC.

Kim et al. (2017) used a cohort of 698 Korean women with dense breasts (BIRADS c or d) presenting for screening or diagnostic work-up at a single institution. For women having diagnostic work-up (n=506), ROC analysis, measured by AUC measurement was reported. Excellent AUC measurement was reported for both FFDM + DBT and FFDM + ultrasound in women with more dense breasts:

- FFDM + ultrasound was 0.958, and
- FFDM + DBT was 0.921.

FFDM + ultrasound had a statistically significant higher sensitivity compared to FFDM + DBT (97.0% compared to 91.1%, $p=.008$) but specificity and PPV for FFDM + DBT was significantly higher (78.4% compared to 60.4%, $p<.001$). In this study, compared to ultrasound, DBT was likely to return fewer false positive results. Kim et al. concluded that FFDM + DBT was non-inferior to FFDM + ultrasound for the detection of cancer in women with more dense breasts. While this study included a mixed population including women presenting for screening and those seeking diagnosis, the study did not discuss characteristics of cancers detected and has limited further relevance to consideration about the role of DBT in the assessment centre.

Overall differences in AUC magnitude between reported in this study and Kim et al.'s paper may be due to differences in imaging protocol (eg, 1v-DBT compared to 2v-DBT) or study population (i.e., women in Thibault's study were possibly more likely to have a breast cancer than women in Kim et al.'s study).

Both Kim et al. and Thibault et al. reported slightly increased PPV with adjunctive DBT protocols compared to ultrasound protocols (60.6% and 57% respectively for the adjunctive DBT compared to 47.1% and 53% for DM + ultrasound protocols), suggesting that more women are correctly undergoing biopsy for malignancy (although the base cohort differed as the result from Kim et al. only discussed PPV for its diagnostic participants). Also, reported NPV was slightly

higher in both studies with adjunctive ultrasound (98.2% and 78% for adjunctive ultrasound compared to 96% and 93% for adjunctive DBT).

Analysis of AUC measurement by breast density (BIRADS c and d) was reported by Kim et al. for both screening and diagnostic populations. The reported results showed no difference in AUC measurement for women with heterogeneously dense breasts imaged with FFDM + DBT compared to women imaged with FFDM + ultrasound (0.949 compared to 0.969, $p=.076$) or for women with the most dense breasts (0.842 compared to 0.931, $p=.091$).

Taken together, results from the two prospective studies indicate that DBT may have a place in the assessment of suspicious lesions due to excellent diagnostic accuracy, but results do not discuss clinical pathway implications or suggest that DBT should replace ultrasound in any way at this time.

Retrospective studies for women recalled to assessment

Lee et al. (2016) completed a small retrospective review of 108 women (17 malignancies) with lesions classified as BIRADS 0 at screening mammography to assess comparative diagnostic accuracy between DBT and ultrasound. Using biopsy/surgery as the reference standard, the authors reported that DBT had a higher level of diagnostic accuracy for BIRADS 0 lesions compared to breast ultrasound and that fewer benign biopsies were performed following imaging with DBT compared to ultrasound (50% for DBT compared to 71.2% for ultrasound) in women with more dense breasts (BIRADS c or d). It is presumed that the DBT and ultrasound imaging were follow-on imaging from screening mammography only (not adjunctive to other mammographic work-up). Specificity and PPV were significantly higher with DBT compared to ultrasound.

Retrospective studies in mixed populations or with an unclear diagnostic pathway

Two retrospective studies reported on the diagnostic accuracy of DM+ DBT compared to DM + ultrasound in women recalled to assessment. Retrospective observational studies report similar results to the prospective studies: adjunctive DBT is not inferior to adjunctive ultrasound in terms of diagnostic accuracy, but superiority of DBT over ultrasound during work-up was neither the focus of the research, nor a clear outcome.

Elizalde et al. (2016) used a cancer-enriched study population of 1041 women and one man with BIRADS 2,3 or 4 breast densities or attending for diagnostic work-up for a screen-detected abnormality (no data was provided on the numbers in each study sub-population). This study aimed to assess the diagnostic accuracy of FFDM + ultrasound to FFDM + DBT_{MLO +/- CC} (among other protocols). Most participants had 1v-DBT where this provided sufficient diagnostic information (only 56 required both MLO and CC views). Elizalde et al. reported, following one radiologist's blinded review of images, that both FFDM + DBT (in one or two views, not stated) and FFDM + ultrasound increased diagnostic accuracy compared to FFDM, with a limited difference in AUC reported (0.85 for FFDM + ultrasound compared to 0.84 for DBT). Both adjunctive modalities increased FFDM sensitivity by 23% ($p<.05$) compared to an increase of 17.8% with FFDM + DBT ($p<.05$). Specificity of FFDM + DBT was statistically higher than FFDM + ultrasound (83.5 compared to 74.32%), which reflects the higher number of false positives seen with ultrasound imaging (usually detection of fibroadenomas and other benign lesions). The authors concluded that FFDM + DBT and FFDM + ultrasound have similar diagnostic accuracy and FFDM + DBT is not inferior but no further comment was made about preferred work-up pathway by mammographic finding/suspected abnormality.

Kim et al. (2015) undertook a retrospective reader performance study with blinding comparing the diagnostic performance of DBT and bilateral whole breast ultrasound in a cancer-enriched

cohort of 113 Korean women with non-calcified lesions identified during screening or diagnostic work-up (75 breast cancers). In the study group, 60% of participants presented with palpable abnormalities and the remainder were asymptomatic. Readers were given a matched set of FFDM + DBT and FFDM + ultrasound images and asked to distinguish between benign and malignant lesions. AUC measurement reported for FFDM + DBT was 0.899 compared to 0.914 for FFDM + ultrasound (although statistical significance was not achieved). The authors also reported that false positive rates of 55% and 60% respectively: five cancers were misclassified on DBT but correctly classified with ultrasound; two cancers were correctly classified as malignant on DBT but were missed not on ultrasound. The authors concluded that performance between the two modalities was comparable and that DBT could be used instead of ultrasound in the characterisation of lesions.

Differences were seen in the types of additional tumours detected with ultrasound or DBT. Although reporting on small numbers (n=20 additional cancers detected with ultrasound and 15 detected with DBT), Elizalde et al. reported that ultrasound detected two more invasive ductal carcinoma (IDC), an ILC and one more tubular carcinoma. No statistical testing was completed.

Table 17: AUC measurement, sensitivity, specificity and PPV/NPV results for adjunctive DBT and adjunctive ultrasound

17A Prospective studies in women presenting for screening and symptomatic women

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Kim et al. (2017)	Prospective non-inferiority reader (n=12) study comparing adjunctive DBT to adjunctive ultrasound Participants: 698 women with dense breasts (192 for screening; 506 for diagnostic work-up), BIRADS c 77.1%, d 22.9%, 140 malignancies Unit: Hologic Selenia Dimensions	<i>All participants</i> FFDM + DBT: 0.933 (0.912, 0.950) FFDM + ultrasound: 0.964 (0.947, 0.976) Difference: -0.031, $p=.014$ <i>For diagnostic exams</i> FFDM + DBT: 0.921 (0.894, 0.943) FFDM + ultrasound: 0.958 (0.937, 0.974) Difference: -0.037, $p=.005$	<i>All participants</i> FFDM + DBT: 91.4 (86.8, 96.0) FFDM + ultrasound: 96.4 (93.3, 99.5), $p=.039$ <i>For diagnostic exams</i> FFDM + DBT: 91.1 (86.3, 95.9) FFDM + ultrasound: 97.0 (94.2, 99.9), $p=.008$	<i>All participants</i> FFDM + DBT: 83.9 (80.9, 86.9) FFDM + ultrasound: 70.4 (66.6, 74.2), $p<.001$ <i>For diagnostic exams</i> FFDM + DBT: 78.4 (74.3, 82.6) FFDM + ultrasound: 60.4 (55.4, 65.4), $p<.001$	<i>PPV for diagnostic exams</i> FFDM + DBT: 60.6 (53.9, 67.3) FFDM + ultrasound: 47.1 (41.2, 53), $p<.001$ <i>NPV for diagnostic exams</i> FFDM + DBT: 96.0 (93.8, 98.2) FFDM + ultrasound: 98.2 (96.5, 99.9), $p=.019$

17B Prospective studies in women recalled to assessment and symptomatic women

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Thibault et al. (2013)	Prospective multisite (n=5) reader (n=7) study comparing DBT _{MLO} to a range of imaging alternatives including FFDM + additional mammographic views and ultrasound Participants: 131 consecutive women recalled to assessment or symptomatic women (55 malignant and 76 benign/normal findings) Unit: adapted Senographe DS platform	FFDM: 0.7556 (0.7185, 0.7927) FFDM + US: 0.7697 (0.7302, 0.8091) FFDM + ultrasound + DBT _{MLO} : 0.7628 (0.7222, 0.8035) DBT _{MLO} alone: 0.7783 (0.7390, 0.8176) DBT _{MLO} + DM _{CC} : 0.7795 (0.7401, 0.8189)	FFDM: 73% FFDM + US: 81% FFDM + ultrasound + DBT: 81% DBT alone: 66% DBT + DM _{CC} : 68%	FFDM: 53% FFDM + US: 48% FFDM + ultrasound + DBT: 52% DBT alone: 64% DBT + DM _{CC} : 64%	<i>PPV</i> FFDM: 53% FFDM + US: 53% FFDM + ultrasound + DBT: 55% DBT alone: 57% DBT + DM _{CC} : 58% <i>NPV</i> FFDM: 74% FFDM + US: 78% FFDM + ultrasound + DBT: 79% DBT alone: 72% DBT + DM _{CC} : 73%

17C Retrospective studies

Study	Study details	AUC measurement (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV and NPV
Elizalde et al. (2016)	Retrospective reader (n=1) study. Participants: 1041 Spanish women and 1 man with biopsy-proven malignancies (n=84), biopsy-proven benign lesions (n=258), and patients with normal or benign findings (n=700) undergoing imaging with FFDM + DBT _{MLO} +/- CC + ultrasound Unit: Siemens Inspiration	FFDM: 0.70 FFDM + ultrasound: 0.85 FFDM + DBT: 0.84 FFDM + DBT + ultrasound: 0.91 <i>FFDM + ultrasound compared to FFDM + DBT</i> <i>P=.7332</i>	FFDM: 69.05 FFDM + ultrasound: 92.86 FFDM + DBT: 86.90 FFDM + DBT + ultrasound: 98.81	FFDM: 88.2 FFDM + ultrasound: 74.32 FFDM + DBT: 83.50 FFDM + DBT + ultrasound: 74.11	No data
Lee et al. (2016)	Retrospective single institution reader (n=3) study. Cohort: 108 women with dense breasts (BIRADS c or d) and 108 lesions (17 malignancies) with a BIRADS 0 classification following screening. Unit: Hologic Dimensions	No AUC measurement but diagnostic accuracy (TP+TN)/(TP+FP+TN+FN)	<i>BIRADS 0 at mammography</i> DBT: 100 Ultrasound: 100 <i>For asymmetries with microcalcifications</i> DBT: 100 Ultrasound: 100	<i>BIRADS 0 at mammography</i> DBT: 81.3 Ultrasound: 53.9 <i>For asymmetries with microcalcifications</i> DBT: 84.7 Ultrasound: 54.1	<i>PPV BIRADS 0 at mammography</i> DBT: 50 Ultrasound: 28.8 <i>NPV BIRADS 0 at mammography</i> DBT: 100 Ultrasound: 100
Kim et al. (2015)	Retrospective reader (n=3) study in a cancer-enriched cohort. Participants: 113 Korean women with 119 breast lesions (excluded patients had AD, asymmetry, microcalcification without mass); 75 cancers; Unit: Selenia Dimensions	Overall Az value DBT: 0.899 Ultrasound: 0.914 Difference: 0.015 (-0.019, 0.049), <i>p</i> =.394	<i>All 3 readers</i> DBT: 97.3 US 98.7 <i>p</i> =.508	<i>All 3 readers</i> DBT: 44.7 US: 39.4 <i>p</i> =.360	<i>PPV for BIRADS 3-5</i> DBT: 67.8 Ultrasound: 66.2 <i>PPV for BIRADS 3-5</i> DBT: 74.9 Ultrasound: 75.3 No significance testing reported

8.2. MRI

MRI is used in breast cancer work-up to:

- further evaluate suspicious areas that may be hard to see on DM or ultrasound (i.e., an adjunct imaging modality to support diagnosis in difficult cases rather than a routine assessment modality for average-risk women), or
- evaluate the size and extent of cancer following a breast cancer diagnosis and to look for additional breast cancer in the ipsilateral or contralateral breast if breast cancer is confirmed (in treatment path planning).

MRI is not used in the BSA program (although women being treated for breast cancer in Australia may have MRI as part of treatment path planning imaging). MRI's limitations in a screening environment include its cost, access to MRI equipment in some areas, that it requires a longer examination time compared to DBT or DM, it requires a contrast agent which may accumulate in the brain over multiple exposures, and that it is not well-tolerated by all women.

This literature review reports on studies with relevance to breast cancer screening and assessment. To reflect the way that MRI could be used in a breast screening assessment process (rather than in diagnosis or treatment management), this section describes studies investigating adjunctive DBT (such as FFDM + DBT) compared to adjunctive MRI (such as DM + MRI). We want to know if DBT alone can provide sufficient additional diagnostic information to result in a reduction in use of MRI in breast cancer assessment.

Out of scope issues included studies about:

- MRI's role in determining the extent of disease
- the use of MRI to guide treatment decision-making or surveillance of treatment effectiveness, or
- the role that DBT could play in assessing further suspicious lesions identified on MRI only (i.e., second-look DBT compared to second-look ultrasound or DM for lesions identified on MRI only).

None of the systematic reviews or narrative literature reviews compared adjunctive DBT to adjunctive MRI in the assessment of lesions suspicious for breast cancer or described situations where DBT had been assessed as a replacement imaging technique for MRI.

The literature search returned two prospective studies and four retrospective studies which explored the role of adjunctive DBT compared to adjunctive MRI. The studies discussed in this literature review are listed below. Study findings are summarised in *Table 18 (A-B)*.

Systematic reviews and meta-analysis

Systematic reviews: none

Narrative literature reviews: none

RCTs and prospective studies

RCTs: none

Two prospective studies:

Women presenting for screening/women with symptoms: Mariscotti et al. (2014)

Retrospective studies

Four retrospective studies:

Women recalled to assessment/women with symptoms: Roganovic et al. (2015)

Only women with symptoms: Kim et al. (2016)

Diagnostic pathway unclear: Fornvik et al. (2018); Kamal et al. (2016)

Key findings

Six studies investigated adjunctive DBT compared to adjunctive MRI, with two prospective studies commenting on whether DBT could provide sufficient additional diagnostic information to result in a reduction in use of MRI in breast cancer assessment.

In symptomatic women and mixed populations, adjunctive DBT is not inferior to adjunctive MRI for the assessment of lesions suspicious for breast cancer but there is no evidence to suggest that DBT should replace MRI in the work-up of suspicious lesions

No statistically significant gain in sensitivity with adjunctive MRI was reported if a lesion had been imaged with FFDM + DBT + ultrasound: sensitivity with FFDM + ultrasound + MRI was 98.8% compared to 97.7% for FFDM + DBT + ultrasound. Overall, consistent findings were reported in the retrospective observational studies: adjunctive MRI has better diagnostic accuracy (as measured by AUC measurement) than adjunctive DBT but MRI had a lower specificity.

8.2.1. In symptomatic women and mixed populations, adjunctive DBT is not inferior to adjunctive MRI for the assessment of lesions suspicious for breast cancer but there is no evidence to suggest that DBT should replace MRI in the work-up of suspicious lesions

Prospective studies of symptomatic women

The literature review returned one small, single-institution prospective study comparing the diagnostic accuracy of adjunctive DBT to adjunctive MRI (that is, FFDM + DBT compared to FFDM + MRI) in symptomatic women. While measures of diagnostic accuracy were higher for FFDM + MRI protocols, statistical significance was not achieved for any results and adjunctive DBT's inferiority to adjunctive MRI was not confirmed.

Tang et al. (2017) completed a study of 197 symptomatic patients or women with ultrasound-detected lesions (153 breast cancers, 68 were multifocal, bilateral or multicentric) attending a hospital in Shanghai aimed to determine an optimal imaging protocol involving FFDM, DBT and MRI. The authors reported AUC measurement, finding that diagnostic accuracy, sensitivity, specificity and PPV/NPV increased with both FFDM + DBT and FFDM + MRI compared to FFDM, with all AUC measurements for FFDM + DBT and FFDM + MRI both exceeding 0.9 for both readers. FFDM + MRI's superiority over FFDM + DBT was not demonstrated (*p* value for the difference between FFDM + DBT compared to FFDM + MRI was .1262 or .0843 for each reader).

Prospective studies of mixed populations

Mariscotti et al. (2014) imaged 200 Italian women with lesions classified as BIRADS ≥ 3 (i.e., inconclusive or suspicious for breast cancer) and who were recalled to assessment or undergoing diagnostic mammography for a symptomatic presentation. Women who had a histologically-proven cancer (257 breast cancers, 59 were multifocal) and who had the full preoperative imaging protocol (FFDM + ultrasound + DBT + MRI) were included. Of all the imaging modalities, MRI had the highest sensitivity. This was particularly apparent for women with more dense breasts (which is expected given that breast density does not affect the MR image). MRI's sensitivity was not significantly superior to FFDM + DBT + ultrasound (98.8% compared to 97.7%; *p*=1). The authors reported that there was limited additional gain in diagnostic accuracy when MRI was added to the imaging protocol (i.e., FFDM + DBT + ultrasound + MRI). Specificity was also not significantly higher for FFDM + ultrasound + DBT compared to MRI (82.8% compared to 74.2%, *p*=.21). Mariscotti et al. concluded that limited further information to support diagnosis was found with MRI compared to the FFDM + DBT + ultrasound protocol.

Retrospective studies

The literature review returned four studies that compared adjunctive DBT to adjunctive MRI in an assessment setting. Overall, consistent findings were reported in these studies: that is, adjunctive MRI has better diagnostic accuracy (as measured by AUC value and higher overall sensitivity) than adjunctive DBT but a lower specificity.

The most recent study comparing DBT and MRI in the diagnosis of breast cancer was published by Fornvik et al. (2018). This matched comparison study compared the diagnostic performance of MRI in 152 Japanese women with dense breasts (BIRADS c or d), a BIRADS ≥ 3 report (inconclusive) on MRI and who were attending follow-up imaging to investigate suspicious DM, ultrasound or tomography findings. All participants had FFDM + DBT + MRI. A statistically significant or borderline significant increase in AUC measurement was reported for MRI (0.964) compared to DBT for either of the two observers (0.926 or 0.875; *p*=.004, *p*=.052). Lesion-

related outcomes were that MRI detected smaller cancers than DBT, leading to increase diagnostic performance for MRI compared to DBT for the two readers involved in the study.

Kim et al. (2016) also completed a retrospective study comparing MRI and DBT as adjunctive assessment imaging for women with known breast cancers. They reported that adjunctive MRI (i.e., FFDM + ultrasound + MRI) had statistically significant higher diagnostic performance than adjunctive DBT (i.e., FFDM + ultrasound + DBT) (JAFROC FOM analysis 0.978 compared to 0.937 respectively). Both adjunctive DBT and adjunctive MRI results demonstrated excellent diagnostic accuracy.

For all studies, reported sensitivity and specificity was higher for MRI compared to DBT. In Fornvik et al.'s 2018 study, sensitivity for MRI reached 97% compared to about 80% for DBT imaging for each reader. PPV varied by observer (74.6-76.1% with DBT) compared to PPV of 62.5% with MRI. MRI correctly identified more women with breast cancer compared to DBT but also reported more false-positives. Possible explanations were that BIRADS 1 and 2 lesions were excluded from Fornvik et al.'s study, possibly leading to higher sensitivity for DBT. Higher sensitivity but lower PPV was also reported in Kim et al. (2016): adjunctive DBT had lower sensitivity (FFDM + DBT 88.2% compared to 98.9% with FFDM +MRI) but higher specificity and PPV compared to adjunctive MRI. Kim et al. noted that this difference in sensitivity could largely be attributed to the effect of breast density which is not an issue with MRI.

Kamal et al. (2016) also reported on the impact of FFDM + DBT or FFDM + MRI on the assessment of breast cancer. In this study, 98 Egyptian women with either a palpable lump and dense breasts (BIRADS c or d) or lesions classified as BIRADS 3 or 4 on FFDM had additional DM, DBT and MRI. There were 25 breast cancers/ As with other studies, the authors reported that both DBT and MRI resulted in increased sensitivity compared to FFDM. Further, a statistically significant increase in sensitivity when using MRI compared to DBT was also reported ($p=.035$); however, this study also reported higher specificity with MRI. Mercier et al. (2015) also reported higher sensitivity with MRI compared to DBT (although limited further comparison information about DBT and MRI findings are provided in Mercier et al.'s paper).

Different findings are also presented by Roganovic et al. (2015), who completed a study to investigate the sensitivity of MRI and DBT compared to histopathology of 57 breast lesions (29 breast cancers). Roganovic et al. reported that DBT had both higher sensitivity and specificity than MRI (100% compared to 93.1% and 75% compared to 60.7%, $p=.02$), higher PPV and NPV compared to MRI, and a larger AUC measurement compared to MRI (0.925 compared to 0.884). These results differ from other studies. One of the reasons that could explain these results was either higher background enhancement with MRI, or that intramammary lymph nodes could appear malignant on MRI, contributing to a higher number of false positives with this imaging modality but they were easily distinguished as benign on DBT.

Table 18: AUC measurement, sensitivity, specificity and PPV/NPV results comparing adjunctive DBT to adjunctive MRI

18A Prospective studies

Study	Study details	AUC measurement (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV and NPV
Tang et al. (2017)	Prospective reader (n=2) study Participants: 197 consecutive Chinese women with 238 suspected lesions (153 malignancies; 85 benign) Unit: Hologic Selenia Dimensions	<i>Reader 1</i> FFDM: 0.849 FFDM + DBT: 0.907 FFDM + MRI: 0.939 $p=.1262$ <i>Reader 2</i> FFDM: 0.850 FFDM + DBT: 0.900 FFDM + MRI: 0.935 $p=.0843$	<i>Reader 1</i> FFDM: 71.9 FFDM + DBT: 92.1 FFDM + MRI: 94.7 <i>Reader 2:</i> FFDM: 71.2 FFDM + DBT: 90.8 FFDM + MRI: 95.4 <i>No significant results for FFDM + DBT compared to FFDM + MRI</i>	<i>Reader 1</i> FFDM: 71.7 FFDM + DBT: 73.0 FFDM + MRI: 82.3 <i>Reader 2</i> FFDM: 71.7 FFDM + DBT: 68.2 FFDM + MRI: 82.3 Difference between FFDM + DBT/FFDM + MRI for Reader 2: $p=.025$	<i>PPV</i> FFDM: 81.9 FFDM + DBT: 83.9 FFDM + MRI: 90.1 <i>NPV</i> FFDM: 58.1 FFDM + DBT: 82.8 FFDM + MRI: 89.7
Mariscotti et al. (2014)	Prospective study Participants: 200 women who had DM + DBT + ultrasound + MRI and who had histology proven breast cancer; 257 malignancies, 93 benign lesions Unit: Hologic Selenia Dimensions	No data	<i>All lesions</i> FFDM: 85.2 (80.5, 89.2) DBT: 90.7 (86.6, 93.8) FFDM + DBT + US: 97.7 (95.0, 98.9) MRI: 98.8 (96.9, 99.7) <i>Difference between DBT protocol and MRI: $p=1$</i>	<i>All lesions</i> FFDM: 86.9 (83.0, 90.1) DBT: 90.0 (86.5, 92.8) FFDM + DBT + US: 93.7, (90.7, 95.8) MRI: 92.3 (89.1, 94.8) <i>Difference between DBT protocol and MRI: $p=.21$</i>	<i>PPV All lesions</i> FFDM: 96.5 (93.2, 98.2) DBT: 95.5 (92.1, 97.5) FFDM + DBT + US: 94.0 (90.5, 96.3) MRI: 91.4 (87.5, 94.1) <i>NPV All lesions</i> FFDM: 69.1 (60.5, 76.6) DBT: 77.4 (68.5, 84.3) FFDM + DBT + US: 96.3 (89.6, 98.7) MRI: 95.8 (88.5, 98.6) <i>PPV difference for DBT protocol and MRI: $p=.88$</i> <i>NPV Difference for DBT protocol and MRI: $p=.76$</i>

18B Retrospective studies

Study	Study details	AUC measurement (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV and NPV
Fornvik et al. (2018)	Retrospective observer (n=2) review Participants: 152 Japanese women with 103 malignancies (64 NST, 26 DCIS, 6 ILC, 4	<i>Parametric ROC DBT</i> Reader 1: 0.875 (0.801, 0.927) Reader 2: 0.906 (0.852, 0.944)	<i>DBT</i> Reader 1: 80.6 Reader 2: 82.5 <i>MRI</i> 97.1%	<i>DBT</i> Reader 1: 80.6 Reader 2: 82.5 <i>MRI</i> Not provided	<i>PPV DBT</i> Reader 1: 76.1 Reader 2: 74.6 <i>PPV MRI</i> 62.5

Study	Study details	AUC measurement (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)	PPV and NPV
	mucinous carcinomas, 3 others) with 3T-MRI and FFDM + DBT in women with more dense breasts (c=67.8%; d=14.5%) Unit: Hologic Selenia Dimensions	<i>Parametric ROC MRI</i> 0.964 (0.931, 0.983), $p=.004$, $p=.052$ (for Readers 1 and 2)			
Kamal et al. (2016)	Retrospective analysis completed by two readers Participants: 98 symptomatic patients with 103 lesions (BIRADS 3 or 4; of which 78 were benign, 25 malignant) assessed with DM, DBT and MR mammography Unit: Hologic Selenia Dimensions	No data	DM: 88 DBT: 92 MRM: 92	DM: 52.56 DBT: 80.77 MRM: 89.74	<i>PPV</i> DM: 37.29 DBT: 60.53 MRM: 74.19 <i>NPV</i> DM: 93.18 DBT: 96.92 MRM: 97.22
Kim et al. (2016)	Retrospective comparison analysis using a prospective cohort and three readers Participants: 172 Korean women with 184 cancers (153 IDC, 15 DCIS, 5 mucinous carcinoma, 4 ILC, 7 others) Unit: Hologic Selenia Dimensions	<i>Pooled JAFROC FOM from three readers</i> FFDM: 0.900 FFDM + DBT: 0.937 FFDM + MRI: 0.978 <i>Difference FFDM/FFDM + DBT: $p=.001$</i> <i>Difference FFDM + DBT/FFDM + MRI: $p=.0006$</i>	FFDM: 78.3 FFDM + DBT: 88.2 FFDM + MRI: 98.9 <i>Difference FFDM/FFDM + DBT: $p=.0001$</i> <i>Difference FFDM + DBT/FFDM + MRI: $p=.0001$</i>	FFDM: 94.5 FFDM + DBT: 93.3 FFDM + MRI: 89.6 <i>Difference FFDM/FFDM + DBT: $p=.2978$</i> <i>Difference FFDM + DBT/FFDM + MRI: $p=.0282$</i>	<i>PPV</i> FFDM + DBT: 93.3 FFDM + MRI: 89.6
Roganovic et al. (2015)	Prospective non-blinded comparison study of three diagnostic imaging modalities: DBT, MRI and DM Participants: 57 breast lesions including 20 IDC, 2 ILC, 7 IDC + DCIS, and 28 benign lesions Unit: Hologic Selenia Dimensions	FFDM: 0.64 DBT: 0.925 MRI: 0.884 Difference: FFDM/DBT: 0.285, $p<.001$ DBT/MRI: 0.041, $p=.002$	FFDM: 72.4% DBT: 100% MRI: 93.1%	FFDM: 46.4% DBT: 75% MRI: 60.7%	<i>PPV</i> : FFDM: 58.3% DBT: 80.6% MRI: 71.1% <i>NPV</i> FFDM: 61.9% DBT: 100% MRI: 89.5%

REFERENCES

- Alakhras M.M., Brennan P.C., Rickard M., Bourne R., & Mello-Thoms C. (2014). Effect of radiologists' experience on breast cancer detection and localization using digital breast tomosynthesis. *European Radiology*, 25(2), 402–409. <https://doi.org/10.1007/s00330-014-3409-1>
- Amer H.A., Schmitzberger F., Ingold-Heppner B., Kussmaul J., El Tohamy M.F., Tantawy H.I., ... Fallenberg E.M. (2017). Digital breast tomosynthesis versus full-field digital mammography-Which modality provides more accurate prediction of margin status in specimen radiography? *European Journal of Radiology*, 93((Amer, El Tohamy, Tantawy) Dept of Radiology, Zagazig University Hospitals, Zagazig, Egypt), 258–264. <https://doi.org/10.1016/j.ejrad.2017.05.041>
- Ariaratnam N.S., Little S.T., Whitley M.A., & Ferguson K. (2018). Digital breast Tomosynthesis vacuum assisted biopsy for Tomosynthesis-detected Sonographically occult lesions. *Clinical Imaging*, 47((Ariaratnam, Little, Whitley, Ferguson) South Jersey Radiology Associates, 1307 White Horse Road, Suite A102, Voorhees, NJ 08043, United States), 4–8. <https://doi.org/10.1016/j.clinimag.2017.08.002>
- Bahl, M., Gaffney, S., McCarthy, A. M., Lowry, K. P., Dang, P. A., & Lehman, C. D. (2018). Breast Cancer Characteristics Associated with 2D Digital Mammography versus Digital Breast Tomosynthesis for Screening-detected and Interval Cancers. *Radiology*, 287(1), 49–57. <https://doi.org/10.1148/radiol.2017171148>
- Bahrs S.D., Otto V., Hattermann V., Klumpp B., Hahn M., Nikolaou K., & Siegmann-Luz K. (2018). Breast tomosynthesis for the clarification of mammographic BI-RADS 3 lesions can decrease follow-up examinations and enables immediate cancer diagnosis. *Acta Radiologica*, ((Bahrs, Otto, Hattermann, Klumpp, Nikolaou, Siegmann-Luz) Department of Diagnostic and Interventional Radiology, University Hospital Tuebingen, Tuebingen, Germany). <https://doi.org/10.1177/0284185118756458>
- Balleyguier C., Arfi-Rouche J., Levy L., Toubiana P.R., Cohen-Scali F., Toledano A.Y., & Boyer B. (2017). Improving digital breast tomosynthesis reading time: A pilot multi-reader, multi-case study using concurrent Computer-Aided Detection (CAD). *European Journal of Radiology*, 97 (Balleyguier, Arfi-Rouche) Department of Radiology, Gustave Roussy, 114 rue Edouard-Vaillant, Villejuif Cedex 94805, France), 83–89. <https://doi.org/10.1016/j.ejrad.2017.10.014>
- Bansal G.J., & Young P. (2015). Digital breast tomosynthesis within a symptomatic 'onestop breast clinic' for characterization of subtle findings. *British Journal of Radiology*, 88(1053), 20140855. <https://doi.org/10.1259/bjr.20140855>
- Berger N., Schwizer S.D., Varga Z., Rageth C., Frauenfelder T., & Boss A. (2016). Assessment of the extent of microcalcifications to predict the size of a ductal carcinoma in situ: comparison between tomosynthesis and conventional mammography. *Clinical Imaging*, 40(6), 1269–1273. <https://doi.org/10.1016/j.clinimag.2016.09.003>
- Bian T., Lin Q., Cui C., Li L., Qi C., Fei J., & Su X. (2016). Digital Breast Tomosynthesis: A New Diagnostic Method for Mass-Like Lesions in Dense Breasts. *Breast Journal*, 22(5), 535–540. <https://doi.org/10.1111/tbj.12622>
- Borelli and Oduko (2016). NHS breast screening programme: current position on use of tomosynthesis.
- Brandt K.R., Craig D.A., Hoskins T.L., Henrichsen T.L., Bendel E.C., Brandt S.R., & Mandrekar J. (2013). Can digital breast tomosynthesis replace conventional diagnostic mammography views

- for screening recalls without calcifications? A comparison study in a simulated clinical setting. *American Journal of Roentgenology*, 200(2), 291–298. <https://doi.org/10.2214/AJR.12.8881>
- Cai, S.-Q., Yan, J.-X., Chen, Q.-S., Huang, M.-L., & Cai, D.-L. (2015). Significance and application of digital breast tomosynthesis for the bi-rads classification of breast cancer. *Asian Pacific Journal of Cancer Prevention*, 16(9), 4109–4114. <https://doi.org/10.7314/APJCP.2015.16.9.4109>
- Caumo, F., Romanucci, G., Hunter, K., Zorzi, M., Brunelli, S., Macaskill, P., & Houssami, N. (2018). Comparison of breast cancers detected in the Verona screening program following transition to digital breast tomosynthesis screening with cancers detected at digital mammography screening. *Breast Cancer Research and Treatment*. <https://doi.org/10.1007/s10549-018-4756-4>
- Chae E.Y., Kim H.H., Cha J.H., Shin H.J., & Choi W.J. (2016). 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. *British Journal of Radiology*, 89(1062), 20150743. <https://doi.org/10.1259/bjr.20150743>
- Chan H.-P., Helvie M.A., Hadjiiski L., Jeffries D.O., Klein K.A., Neal C.H., ... Roubidoux M.A. (2017). Characterization of Breast Masses in Digital Breast Tomosynthesis and Digital Mammograms: An Observer Performance Study. *Academic Radiology*, 24(11), 1372–1379. <https://doi.org/10.1016/j.acra.2017.04.016>
- Choi, J. S., Han, B.-K., Ko, E. Y., Ko, E. S., Hahn, S. Y., Shin, J. H., & Kim, M. J. (2016a). Comparison between two-dimensional synthetic mammography reconstructed from digital breast tomosynthesis and full-field digital mammography for the detection of T1 breast cancer. *European Radiology*, 26(8), 2538–2546. <https://doi.org/10.1007/s00330-015-4083-7>
- Choi W.J., Kim H.H., Lee S.Y., Chae E.Y., Shin H.J., Cha J.H., ... Choi Y.-W. (2016b). A comparison between digital breast tomosynthesis and full-field digital mammography for the detection of breast cancers. *Breast Cancer*, 23(6), 886–892. <https://doi.org/10.1007/s12282-015-0656-1>
- Clauser, P., Nagl, G., Helbich, T. H., Pinker-Domenig, K., Weber, M., Kapetas, P., ... Baltzer, P. A. T. (2016). Diagnostic performance of digital breast tomosynthesis with a wide scan angle compared to full-field digital mammography for the detection and characterization of microcalcifications. *European Journal of Radiology*, 85(12), 2161–2168. <https://doi.org/10.1016/j.ejrad.2016.10.004>
- Conant, E. F. (2014). Clinical Implementation of Digital Breast Tomosynthesis. *Radiologic Clinics of North America*, 52(3), 499–518. <https://doi.org/10.1016/j.rcl.2013.11.013>
- Cornford E.J., Turnbull A.E., James J.J., Tsang R., Akram T., Burrell H.C., ... Jones V. (2016). Accuracy of GE digital breast tomosynthesis vs supplementary mammographic views for diagnosis of screen-detected soft-tissue breast lesions. *British Journal of Radiology*, 89(1058), 20150735. <https://doi.org/10.1259/bjr.20150735>
- Department of Health and Ageing. (2009). *BreastScreen Australia Position Statement on DBT*
- Destounis S. (2018). Role of Digital Breast Tomosynthesis in Screening and Diagnostic Breast Imaging. *Seminars in Ultrasound, CT and MRI*, ((Destounis) Elizabeth Wende Breast Care, LLC, Rochester, NY). <https://doi.org/10.1053/j.sult.2017.08.002>
- Destounis S., Arieno A., & Morgan R. (2013). Preliminary clinical experience with digital breast tomosynthesis in the visualization of breast microcalcifications. *Journal of Clinical Imaging Science*, 3(1), 65. <https://doi.org/10.4103/2156-7514.124099>
- Destounis S., Santacroce A., & Arieno A. (2017). DBT as a Screening Tool and a Diagnostic Tool. *Current Breast Cancer Reports*, 9(4), 264–271. <https://doi.org/10.1007/s12609-017-0259-2>

- Dibble, E. H., Lourenco, A. P., Baird, G. L., Ward, R. C., Maynard, A. S., & Mainiero, M. B. (2018). Comparison of digital mammography and digital breast tomosynthesis in the detection of architectural distortion. *European Radiology*, 28(1), 3–10. <https://doi.org/10.1007/s00330-017-4968-8>
- Eghtedari M., Tsai C., Robles J., Blair S.L., & Ojeda-Fournier H. (2018). Tomosynthesis in Breast Cancer Imaging: How Does It Fit into Preoperative Evaluation and Surveillance? *Surgical Oncology Clinics of North America*, 27(1), 33–49. <https://doi.org/10.1016/j.soc.2017.07.004>
- Elizalde A., Pina L., Etxano J., Slon P., Zalazar R., & Caballeros M. (2016). Additional US or DBT after digital mammography: Which one is the best combination? *Acta Radiologica*, 57(1), 13–18. <https://doi.org/10.1177/0284185114563641>
- Endo T., Morita T., Oiwa M., Suda N., Sato Y., Ichihara S., ... Arai T. (2018). Diagnostic performance of digital breast tomosynthesis and full-field digital mammography with new reconstruction and new processing for dose reduction. *Breast Cancer*, 25(2), 159–166. <https://doi.org/10.1007/s12282-017-0805-9>
- Fornvik D., Kataoka M., Iima M., Ohashi A., Kanao S., Toi M., & Togashi K. (2018). The role of breast tomosynthesis in a predominantly dense breast population at a tertiary breast centre: breast density assessment and diagnostic performance in comparison with MRI. *European Radiology*, ((Fornvik) Medical Radiation Physics, Department of Translational Medicine, Faculty of Medicine, Lund University, Skane University Hospital, Malmö 205 02, Sweden), 1–10. <https://doi.org/10.1007/s00330-017-5297-7>
- Fornvik D., Zackrisson S., Ljungberg O., Svahn T., Timberg P., Tingberg A., & Andersson I. (2010). Breast tomosynthesis: Accuracy of tumor measurement compared with digital mammography and ultrasonography. *Acta Radiologica (Stockholm, Sweden: 1987)*, 51(3), 240–247. <https://doi.org/10.3109/02841850903524447>
- Freer P.E., Niell B., & Rafferty E.A. (2015). Preoperative tomosynthesis-guided needle localization of mammographically and sonographically occult breast lesions. *Radiology*, 275(2), 377–383. <https://doi.org/10.1148/radiol.14140515>
- Galati F., Marzocca F., Bassetti E., Luciani M.L., Tan S., Catalano C., & Pediconi F. (2017). Added value of digital breast tomosynthesis combined with digital mammography according to reader agreement: Changes in bi-rads rate and follow-up management. *Breast Care*, 12(4), 218–222. <https://doi.org/10.1159/000477537>
- Garayoa J., Chevalier, M., Castillo, M., Mahillo-Fernández, I., Amallal El Ouahabi, N., Estrada, C., ... Valverde, J. (2018). Diagnostic value of the stand-alone synthetic image in digital breast tomosynthesis examinations. *European Radiology*, 28(2), 565–572. <https://doi.org/10.1007/s00330-017-4991-9>
- Garayoa J., Hernández-Girón I., Valverde Morán J. (2014). Breast tomosynthesis: Image quality and Dose. Breast Imaging: 12th International Workshop, IWDM 2014. Fujita Takeshi H, Chisako M Meds. Gifu City, Japan.
- Garcia-Leon, F. J., Llanos-Mendez, A., & Isabel-Gomez, R. (2015). Digital tomosynthesis in breast cancer: A systematic review. *Radiologica*, 57(4), 333–343. <https://doi.org/10.1016/j.rx.2014.06.006>
- Gartner Roth, R., Maidment, A. D. A., Weinstein, S. P., Orel Roth, S., & Conant, E. F. (2014). Digital breast tomosynthesis: Lessons learned from early clinical implementation. *Radiographics*, 34(4), E89–E102. <https://doi.org/10.1148/rg.344130087>

- Gaur, S., Dialani, V., Slanetz, P. J., & Eisenberg, R. L. (2013). Architectural Distortion of the Breast. *American Journal of Roentgenology*, 201(5), W662–W670. <https://doi.org/10.2214/AJR.12.10153>
- Gennaro, G., Hendrick, R. E., Toledano, A., Paquetet, J. R., Bezzon, E., Chersevani, R., ... Muzzio, P. C. (2013a). Combination of one-view digital breast tomosynthesis with one-view digital mammography versus standard two-view digital mammography: per lesion analysis. *European Radiology*, 23(8), 2087–2094. <https://doi.org/10.1007/s00330-013-2831-0>
- Gennaro G., Hendrick R.E., Ruppel P., Chersevani R., Di Maggio C., La Grassa M., ... Muzzio P.C. (2013b). Performance comparison of single-view digital breast tomosynthesis plus single-view digital mammography with two-view digital mammography. *European Radiology*, 23(3), 664–672. <https://doi.org/10.1007/s00330-012-2649-1>
- Gilbert F.J., Tucker L., Gillan M.G.C., Willsher P., Cooke J., Duncan K.A., ... Duffy S.W. (2015a). Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (tommy trial)1. *Radiology*, 277(3), 697–706. <https://doi.org/10.1148/radiol.2015142566>
- Gilbert F.J., Tucker L., Gillan M.G.C., Willsher P., Cooke J., Duncan K.A., ... Duffy S.W. (2015b). 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 Technology Assessment*, 19(4), 1–166. <https://doi.org/10.3310/hta19040>
- Hakim, C. M., Chough, D. M., Ganott, M. A., Sumkin, J. H., Zuley, M. L., & Gur, D. (2010). Digital Breast Tomosynthesis in the Diagnostic Environment: A Subjective Side-by-Side Review. *American Journal of Roentgenology*, 195(2), W172–W176. <https://doi.org/10.2214/AJR.09.3244>
- Haq R., Lim Y.Y., Maxwell A.J., Hurley E., Beetles U., Bundred S., ... Gilbert F.J. (2015). Digital breast tomosynthesis at screening assessment: Are two views always necessary? *British Journal of Radiology*, 88(1055). <https://doi.org/10.1259/bjr.20150353>
- Helal M.H., Mansour S.M., Zaglol M., Salaleldin L.A., Nada O.M., & Haggag M.A. (2017). Staging of breast cancer and the advanced applications of digital mammogram: What the physician needs to know? *British Journal of Radiology*, 90(1071), 20160717. <https://doi.org/10.1259/bjr.20160717>
- Heywang-Kobrunner S.H., Hacker A., Jansch A., Kates R., & Wulz-Horber S. (2017). Use of single-view digital breast tomosynthesis (DBT) and ultrasound vs. additional views and ultrasound for the assessment of screen-detected abnormalities: German multi-reader study. *Acta Radiologica*, ((Heywang-Kobrunner, Hacker, Jansch, Wulz-Horber) National Reference Centre Mammography Munich, Munich, Germany). <https://doi.org/10.1177/0284185117732600>
- Hodgson, R., Heywang-Köbrunner, S. H., Harvey, S. C., Edwards, M., Shaikh, J., Arber, M., & Glanville, J. (2016). Systematic review of 3D mammography for breast cancer screening. *Breast (Edinburgh, Scotland)*, 27(9213011), 52–61. <https://doi.org/10.1016/j.breast.2016.01.002> (Original work published)
- Houssami N., Lång K., Bernardi D., Tagliafico A., Zackrisson S., & Skaane P. (2016). Digital breast tomosynthesis (3D-mammography) screening: A pictorial review of screen-detected cancers and false recalls attributed to tomosynthesis in prospective screening trials. *Breast*, 26((Houssami) School of Public Health (A27), Sydney Medical School, University of Sydney, Sydney 2006, Australia), 119–134. <https://doi.org/10.1016/j.breast.2016.01.007>

Johnson M.M. (2017). Full-Field Digital Mammography and Digital Breast Tomosynthesis. *Radiologic Technology*, 88(3), 299M-319M.

Kamal R., Mansour S., ElMesidy D., Moussa K., & Hussien A. (2016). Detection and diagnosis of breast lesions: Performance evaluation of digital breast tomosynthesis and magnetic resonance mammography. *Egyptian Journal of Radiology and Nuclear Medicine*, 47(3), 1159–1172. <https://doi.org/10.1016/j.ejrn.2016.06.008>

Kang, H.-J., Chang, J. M., Lee, J., Song, S. E., Shin, S. U., Kim, W. H., ... Moon, W. K. (2016). Replacing single-view mediolateral oblique (MLO) digital mammography (DM) with synthesized mammography (SM) with digital breast tomosynthesis (DBT) images: Comparison of the diagnostic performance and radiation dose with two-view DM with or without MLO-DBT. *European Journal of Radiology*, 85(11), 2042–2048. <https://doi.org/10.1016/j.ejrad.2016.09.007>

Kim S.A., Chang J.M., Cho N., Yi A., & Moon W.K. (2015). Characterization of breast lesions: Comparison of digital breast tomosynthesis and ultrasonography. *Korean Journal of Radiology*, 16(2), 229–238. <https://doi.org/10.3348/kjr.2015.16.2.229>

Kim, W. H., Chang, J. M., Lee, J., Chu, A. J., Seo, M., Gweon, H. M., ... Moon, W. K. (2017). Diagnostic performance of tomosynthesis and breast ultrasonography in women with dense breasts: a prospective comparison study. *Breast Cancer Research and Treatment*, 162(1), 85–94. <https://doi.org/10.1007/s10549-017-4105-z>

Kim, W. H., Chang, J. M., Moon, H.-G., Yi, A., Koo, H. R., Gweon, H. M., & Moon, W. K. (2016). Comparison of the diagnostic performance of digital breast tomosynthesis and magnetic resonance imaging added to digital mammography in women with known breast cancers. *European Radiology*, 26(6), 1556–1564. <https://doi.org/10.1007/s00330-015-3998-3>

Kopans, D., Gavenonis, S., Halpern, E., & Moore, R. (2011). Calcifications in the breast and digital breast tomosynthesis. *Breast Journal*, 17(6), 638–644. <https://doi.org/10.1111/j.1524-4741.2011.01152.x>

Lång, K., Andersson, I., & Zackrisson, S. (2014). Breast cancer detection in digital breast tomosynthesis and digital mammography-a side-by-side review of discrepant cases. *The British Journal of Radiology*, 87(1040), 20140080. <https://doi.org/10.1259/bjr.20140080>

Lee C.I., & Lehman C.D. (2016). Digital Breast Tomosynthesis and the Challenges of Implementing an Emerging Breast Cancer Screening Technology Into Clinical Practice. *Journal of the American College of Radiology*, 13(11 Supplement), R61–R66. <https://doi.org/10.1016/j.jacr.2016.09.029>

Lee, W. K., Chung, J., Cha, E.-S., Lee, J. E., & Kim, J. H. (2016). Digital breast tomosynthesis and breast ultrasound: Additional roles in dense breasts with category 0 at conventional digital mammography. *European Journal of Radiology*, 85(1), 291–296. <https://doi.org/10.1016/j.ejrad.2015.09.026>

Lei J., Yang P., Zhang L., Wang Y., & Yang K. (2014). Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: A meta-analysis. *European Radiology*, 24(3), 595–602. <https://doi.org/10.1007/s00330-013-3012-x>

Lockie D., Nickson C., & Aitken Z. (2014). Evaluation of digital breast tomosynthesis (DBT) in an Australian BreastScreen assessment service. *Journal of Medical Imaging and Radiation Oncology*, 58(SUPPL. 1), 74. <https://doi.org/10.1111/1754-9485.12220> (Original work published)

- Lourenco, A. P., Barry-Brooks, M., Baird, G. L., Tuttle, A., & Mainiero, M. B. (2014). Changes in Recall Type and Patient Treatment Following Implementation of Screening Digital Breast Tomosynthesis. *Radiology*, 274(2), 337–342. <https://doi.org/10.1148/radiol.14140317>
- Luparia, A., Mariscotti, G., Durando, M., Ciatto, S., Bosco, D., Campanino, P. P., ... Gandini, G. (2013). Accuracy of tumour size assessment in the preoperative staging of breast cancer: comparison of digital mammography, tomosynthesis, ultrasound and MRI. *La Radiologia Medica*, 118(7), 1119–1136. <https://doi.org/10.1007/s11547-013-0941-z>
- Mall, S., Lewis, S., Brennan, P., Noakes, J., & Mello-Thoms, C. (2017). The role of digital breast tomosynthesis in the breast assessment clinic: a review. *Journal of Medical Radiation Sciences*, 64(3), 203–211. <https://doi.org/10.1002/jmrs.230>
- Mansour S., Adel L., Mokhtar O., & Omar O.S. (2014). Comparative study between breast tomosynthesis and classic digital mammography in the evaluation of different breast lesions. *Egyptian Journal of Radiology and Nuclear Medicine*, 45(3), 1053–1061. <https://doi.org/10.1016/j.ejrn.2014.03.007>
- Mariscotti G., Durando M., Houssami N., Fasciano M., Tagliafico A., Bosco D., ... Gandini G. (2017). Comparison of synthetic mammography, reconstructed from digital breast tomosynthesis, and digital mammography: evaluation of lesion conspicuity and BI-RADS assessment categories. *Breast Cancer Research and Treatment*, 166(3), 765–773. <https://doi.org/10.1007/s10549-017-4458-3>
- Mariscotti G., Durando M., Houssami N., Zuiani C., Martincich L., Londero V., ... Gandini G. (2016). Digital breast tomosynthesis as an adjunct to digital mammography for detecting and characterising invasive lobular cancers: a multi-reader study. *Clinical Radiology*, 71(9), 889–895. <https://doi.org/10.1016/j.crad.2016.04.004>
- Mariscotti, G., Houssami, N., Durando, M., Bergamasco, L., Campanino, P. P., Ruggieri, C., ... Gandini, G. (2014). Accuracy of mammography, digital breast tomosynthesis, ultrasound and MR imaging in preoperative assessment of breast cancer. *Anticancer Research*, 34(3), 1219–1225.
- McDonald E.S., McCarthy A.M., Weinstein S.P., Schnall M.D., & Conant E.F. (2017). Bi-rads category 3 comparison: Probably benign category after recall from screening before and after implementation of digital breast tomosynthesis. *Radiology*, 285(3), 778–787. <https://doi.org/10.1148/radiol.2017162837>
- Mercier, J., Kwiatkowski, F., Abrial, C., Boussion, V., Dieu-de Fraissinette, V., Marraoui, W., ... Lemery, S. (2015). The role of tomosynthesis in breast cancer staging in 75 patients. *Diagnostic and Interventional Imaging*, 96(1), 27–35. <https://doi.org/10.1016/j.diii.2014.06.010>
- Michell M.J., & Batohi B. (2018). Role of tomosynthesis in breast imaging going forward. *Clinical Radiology*, 73(4), 358–371. <https://doi.org/10.1016/j.crad.2018.01.001>
- Morel J.C., Iqbal A., Wasan R.K., Peacock C., Evans D.R., Rahim R., ... Michell M.J. (2014). The accuracy of digital breast tomosynthesis compared with coned compression magnification mammography in the assessment of abnormalities found on mammography. *Clinical Radiology*, 69(11), 1112–1116. <https://doi.org/10.1016/j.crad.2014.06.005>
- Moseley, T. W. (2016). Digital mammography and digital breast tomosynthesis. *Clinical Obstetrics and Gynecology*, 59(2), 362–379. <https://doi.org/10.1097/GRE.0000000000000206>
- Munir A., Huws A., Moalla A., Udayasankar S., Williams H., Lodwich E., ... Holt S. (2016). Our initial experience of digital breast tomosynthesis guided vacuum assisted breast biopsies and

the patient's perspective: A single centre experience. *European Journal of Surgical Oncology*, 42(5), S11. <https://doi.org/10.1016/j.ejso.2016.02.051>

Nakashima, K., Uematsu, T., Itoh, T., Takahashi, K., Nishimura, S., Hayashi, T., & Sugino, T. (2017). Comparison of visibility of circumscribed masses on Digital Breast Tomosynthesis (DBT) and 2D mammography: are circumscribed masses better visualized and assured of being benign on DBT? *European Radiology*, 27(2), 570–577. <https://doi.org/10.1007/s00330-016-4420-5>

Nam, K. J., Han, B.-K., Ko, E. S., Choi, J. S., Ko, E. Y., Jeong, D. W., & Choo, K. S. (2015). Comparison of full-field digital mammography and digital breast tomosynthesis in ultrasonography-detected breast cancers. *Breast (Edinburgh, Scotland)*, 24(5), 649–655. <https://doi.org/10.1016/j.breast.2015.07.039>

Neville, A. M., Bettelheim, R., Gelber, R. D., S  ve-S  derbergh, J., Davis, B. W., Reed, R., ... Price, K. N. (1992). Factors predicting treatment responsiveness and prognosis in node-negative breast cancer. The International (Ludwig) Breast Cancer Study Group. *Journal of Clinical Oncology*, 10(5), 696–705. <https://doi.org/10.1200/JCO.1992.10.5.696>

Ni Mhuirheartaigh N., Coffey L., Fleming H., O' Doherty A., & McNally S. (2017). With the Advent of Tomosynthesis in the Workup of Mammographic Abnormality, is Spot Compression Mammography Now Obsolete? An Initial Clinical Experience. *Breast Journal*, 23(5), 509–518. <https://doi.org/10.1111/tbj.12787>

Noroozian M., Hadjiiski L., Rahnama-Moghadam S., Klein K.A., Jeffries D.O., Pinsky R.W., ... Roubidoux M.A. (2012). Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology*, 262(1), 61–68. <https://doi.org/10.1148/radiol.11101763>

Partyka, L., Lourenco, A. P., & Mainiero, M. B. (2014). Detection of Mammographically Occult Architectural Distortion on Digital Breast Tomosynthesis Screening: Initial Clinical Experience. *American Journal of Roentgenology*, 203(1), 216–222. <https://doi.org/10.2214/AJR.13.11047>

Odle T.G. (2016). Digital Breast Imaging Workflow. *Radiologic Technology*, 87(3), 305M-326M.

Peppard H.R., Nicholson B.E., Rochman C.M., Merchant J.K., Mayo R.C., & Harvey J.A. (2015). Digital Breast Tomosynthesis in the Diagnostic Setting: Indications and Clinical Applications. *Radiographics : A Review Publication of the Radiological Society of North America, Inc*, 35(4), 975–990. <https://doi.org/10.1148/rg.2015140204>

Phi X.-A., Tagliafico A., Houssami N., Greuter M.J.W., & de Bock G.H. (2018). Digital breast tomosynthesis for breast cancer screening and diagnosis in women with dense breasts - a systematic review and meta-analysis. *BMC Cancer*, 18(1), 380. <https://doi.org/10.1186/s12885-018-4263-3>

Rafferty, E. A., Park, J. M., Philpotts, L. E., Poplack, S. P., Sumkin, J. H., Halpern, E. F., & Niklason, L. T. (2013). Assessing Radiologist Performance Using Combined Digital Mammography and Breast Tomosynthesis Compared with Digital Mammography Alone: Results of a Multicenter, Multireader Trial. *Radiology*, 266(1), 104–113. <https://doi.org/10.1148/radiol.12120674>

Rafferty E.A., Park J.M., Philpotts L.E., Poplack S.P., Sumkin J.H., Halpern E.F., & Niklason L.T. (2014). Diagnostic accuracy and recall rates for digital mammography and digital mammography combined with one-view and two-view tomosynthesis: Results of an enriched reader study. *American Journal of Roentgenology*, 202(2), 273–281. <https://doi.org/10.2214/AJR.13.11240>

- Raghu M., Durand M.A., Andrejeva L., Goehler A., Michalski M.H., Geisel J.L., ... Philpotts L.E. (2016). Tomosynthesis in the diagnostic setting: Changing rates of BI-RADS final assessment over time. *Radiology*, 281(1), 54–61. <https://doi.org/10.1148/radiol.2016151999>
- Ray, K. M., Turner, E., Sickles, E. A., & Joe, B. N. (2015). Suspicious Findings at Digital Breast Tomosynthesis Occult to Conventional Digital Mammography: Imaging Features and Pathology Findings. *The Breast Journal*, 21(5), 538–542. <https://doi.org/10.1111/tbj.12446>
- Refaat R., & Matar M.M. (2015). Digital breast tomosynthesis compared to digital mammography in a series of Egyptian women with pathologically proven breast cancer. *Egyptian Journal of Radiology and Nuclear Medicine*, ((Refaat) Department of Radiodiagnosis, Ain Shams University, Cairo, Egypt). <https://doi.org/10.1016/j.ejrm.2015.06.007>
- Roganovic, D., Djilas, D., Vujnovic, S., Pavic, D., & Stojanov, D. (2015). Breast MRI, digital mammography and breast tomosynthesis: comparison of three methods for early detection of breast cancer. *Bosnian Journal of Basic Medical Sciences*, 15(4), 64–68. <https://doi.org/10.17305/bjbms.2015.616>
- Schrading, S., Distelmaier, M., Dirrichs, T., Detering, S., Brolund, L., Strobel, K., & Kuhl, C. K. (2015). Digital breast tomosynthesis-guided vacuum-assisted breast biopsy: initial experiences and comparison with prone stereotactic vacuum-assisted biopsy. *Radiology*, 274(3), 654–662. <https://doi.org/10.1148/radiol.14141397>
- Seo M., Chang J.M., Kim S.A., Kim W.H., Lim J.H., Lee S.H., ... Moon W.K. (2016). Addition of digital breast tomosynthesis to full-field digital mammography in the diagnostic setting: Additional value and cancer detectability. *Journal of Breast Cancer*, 19(4), 438–446. <https://doi.org/10.4048/jbc.2016.19.4.438>
- Seo N., Kim H.H., Shin H.J., Cha J.H., Kim H., Moon J.H., ... Son B.H. (2014). Digital breast tomosynthesis versus full-field digital mammography: Comparison of the accuracy of lesion measurement and characterization using specimens. *Acta Radiologica*, 55(6), 661–667. <https://doi.org/10.1177/0284185113503636>
- Sharma, G. N., Rahul, D., Sanadya, J., Sharma, P., & Sharma, K. K. (2010). Various types and management of breast cancer: an overview. *Journal of Advanced Pharmaceutical Technology & Research*, 1(2), 109–126.
- Skaane P., Gullien R., Bjorndal H., Eben E.B., Ekseth U., Haakenaasen U., ... Krager M. (2012). Digital breast tomosynthesis (DBT): initial experience in a clinical setting. *Acta Radiologica (Stockholm, Sweden : 1987)*, 53(5), 524–529.
- Spangler, M. L., Zuley, M. L., Sumkin, J. H., Abrams, G., Ganott, M. A., Hakim, C., ... Gur, D. (2011). Detection and classification of calcifications on digital breast tomosynthesis and 2D digital mammography: A comparison. *American Journal of Roentgenology*, 196(2), 320–324. <https://doi.org/10.2214/AJR.10.4656>
- Tagliafico, A., Astengo, D., Cavagnetto, F., Rosasco, R., Rescinito, G., Monetti, F., & Calabrese, M. (2012). One-to-one comparison between digital spot compression view and digital breast tomosynthesis. *European Radiology*, 22(3), 539–544. <https://doi.org/10.1007/s00330-011-2305-1>
- Tagliafico A., Gristina L., Bignotti B., Valdora F., Tosto S., & Calabrese M. (2015a). Effects on short-term quality of life of vacuum-assisted breast biopsy: Comparison between digital breast tomosynthesis and digital mammography. *British Journal of Radiology*, 88(1056), 20150593. <https://doi.org/10.1259/bjr.20150593>

- Tagliafico A., Mariscotti G., Durando M., Stevanin C., Tagliafico G., Martino L., ... Houssami N. (2015b). Characterisation of microcalcification clusters on 2D digital mammography (FFDM) and digital breast tomosynthesis (DBT): does DBT underestimate microcalcification clusters? Results of a multicentre study. *European Radiology*, 25(1), 9–14. <https://doi.org/10.1007/s00330-014-3402-8>
- Tagliafico, A. S., Calabrese, M., Bignotti, B., Signori, A., Fiscì, E., Rossi, F., ... Houssami, N. (2017). Accuracy and reading time for six strategies using digital breast tomosynthesis in women with mammographically negative dense breasts. *European Radiology*, 27(12), 5179–5184. <https://doi.org/10.1007/s00330-017-4918-5>
- Taha Ali T.F., Magid A.M.A., Tawab M.A., El-Hariri M.A., & El-Shiekh A.-F. (2016). Potential impact of tomosynthesis on the detection and diagnosis of breast lesions. *Egyptian Journal of Radiology and Nuclear Medicine*, 47(1), 351–361. <https://doi.org/10.1016/j.ejrm.2015.10.006>
- Takamoto Y., Tsunoda H., Kikuchi M., Hayashi N., Honda S., Koyama T., ... Yamauchi H. (2013). Role of breast tomosynthesis in diagnosis of breast cancer for Japanese women. *Asian Pacific Journal of Cancer Prevention*, 14(5), 3037–3040. <https://doi.org/10.7314/APJCP.2013.14.5.303>
- Tang W., Hu F.-X., Zhu H., Wang Q.-F., Gu Y.-J., & Peng W.-J. (2017). Digital breast tomosynthesis plus mammography, magnetic resonance imaging plus mammography and mammography alone: A comparison of diagnostic performance in symptomatic women. *Clinical Hemorheology and Microcirculation*, 66(2), 105–116. <https://doi.org/10.3233/CH-16242>
- Taskin, F., Durum, Y., Soyder, A., & Unsal, A. (2017). Review and management of breast lesions detected with breast tomosynthesis but not visible on mammography and ultrasonography. *Acta Radiologica (Stockholm, Sweden: 1987)*, 58(12), 1442–1447. <https://doi.org/10.1177/0284185117710681>
- Thibault F., Dromain C., Breucq C., Balleyguier C.S., Malhaire C., Steyaert L., ... Drevon H. (2013). Digital breast tomosynthesis versus mammography and breast ultrasound: A multireader performance study. *European Radiology*, 23(9), 2441–2449. <https://doi.org/10.1007/s00330-013-2863-5>
- Thomassin-Naggara, I., Perrot, N., Dechoux, S., Ribeiro, C., Chopier, J., & De Bazelaire, C. (2015). Added value of one-view breast tomosynthesis combined with digital mammography according to reader experience. *European Journal of Radiology*, 84(2), 235–241. <https://doi.org/10.1016/j.ejrad.2014.10.022>
- Tingberg, A., & Zackrisson, S. (2011). Digital mammography and tomosynthesis for breast cancer diagnosis. *Expert Opinion on Medical Diagnostics*, 5(6), 517–526. <https://doi.org/10.1517/17530059.2011.616492>
- Tucker L., Gilbert F.J., Astley S.M., Dibden A., Seth A., Morel J., ... Duffy S.W. (2017). Does reader performance with digital breast tomosynthesis vary according to experience with two-dimensional mammography? *Radiology*, 283(2), 371–380. <https://doi.org/10.1148/radiol.2017151936>
- Waldherr C., Berclaz G., Altermatt H.J., Cerny P., Keller P., Dietz U., ... Sonnenschein M.J. (2016). Tomosynthesis-guided vacuum-assisted breast biopsy: A feasibility study. *European Radiology*, 26(6), 1582–1589. <https://doi.org/10.1007/s00330-015-4009-4>
- Waldherr, C., Cerny, P., Altermatt, H. J., Berclaz, G., Ciriolo, M., Buser, K., & Sonnenschein, M. J. (2013). 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. American Journal of Roentgenology*, 200(1), 226–231. <https://doi.org/10.2214/AJR.11.8202>

- Wallis, M. G., Moa, E., Zanca, F., Leifland, K., & Danielsson, M. (2012). Two-view and single-view tomosynthesis versus full-field digital mammography: High-resolution x-ray imaging observer study. *Radiology*, 262(3), 788–796. <https://doi.org/10.1148/radiol.11103514>
- Wang, W.-S., Hardesty, L., Borgstede, J., Takahashi, J., & Sams, S. (2016). Breast Cancers Found with Digital Breast Tomosynthesis: A Comparison of Pathology and Histologic Grade. *The Breast Journal*, 22(6), 651–656. <https://doi.org/10.1111/tbj.12649>
- Whelehan P., Heywang-Kobrunner S.H., Vinnicombe S.J., Hacker A., Jansch A., Hapca A., ... Evans A. (2017). Clinical performance of Siemens digital breast tomosynthesis versus standard supplementary mammography for the assessment of screen-detected soft-tissue abnormalities: a multi-reader study. *Clinical Radiology*, 72(1), 95. <https://doi.org/10.1016/j.crad.2016.08.011>
- Yang T.-L., Liang H.-L., Chou C.-P., Huang J.-S., & Pan H.-B. (2013). The adjunctive digital breast tomosynthesis in diagnosis of breast cancer. *BioMed Research International*, 2013((Yang, Liang, Chou, Huang, Pan) Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung 81362, Taiwan (Republic of China)), 597253. <https://doi.org/10.1155/2013/597253>
- Yun S.J., Ryu C.-W., Rhee S.J., Ryu J.K., & Oh J.Y. (2017). Benefit of adding digital breast tomosynthesis to digital mammography for breast cancer screening focused on cancer characteristics: a meta-analysis. *Breast Cancer Research and Treatment*, 164(3), 557–569. <https://doi.org/10.1007/s10549-017-4298-1>
- Zuley M.L., Bandos A.I., Ganott M.A., Sumkin J.H., Kelly A.E., Catullo V.J., ... Gur D. (2013). Digital breast tomosynthesis versus supplemental diagnostic mammographic views for evaluation of noncalcified breast lesions. *Radiology*, 266(1), 89–95. <https://doi.org/10.1148/radiol.12120552>

APPENDX A: POPULATIONS INCLUDED IN PRIMARY STUDIES

Only women recalled to assessment	Women recalled to assessment AND women undergoing work-up for a symptom of breast cancer	Women presenting for screening AND women undergoing work-up for a symptom of breast cancer	Only women undergoing work-up for a symptom of breast cancer or with a diagnosed breast cancer	Insufficient detail provided to determine pathway to study recruitment
Bahl et al. (2018) Brandt et al. (2013) Caumo et al. (2018) Choi et al. (2016a) Dibble et al. (2018) Gilbert et al. (2015a,b) Haq et al. (2015) Heywang-Köbrunner et al. (2017) Kopans et al. (2011) Lee et al. (2016) Lockie et al. (unpublished) Lourenco et al. (2014) McDonald et al. (2017) Nam et al. (2015) Ni Mhuircheartaigh et al. (2017) Partyka et al. (2014) Tagliafico et al. (2012) Tucker et al. (2017) Wang et al. (2016) Whelehan et al. (2017)	Balleyguier et al. (2017) Chae et al. (2016) Förnvik et al. (2010) Garayoa et al. (2018) Hakim et al. (2010) Kang et al. (2016) Morel et al. (2014) Rafferty et al. (2013) Ray et al. (2015) Roganovic et al. (2015) Seo et al. (2016) Skaane et al. (2012) Thibault et al. (2013) Waldherr et al. (2013) Wallis et al. (2012)	Galati et al. (2017) Kim et al. (2017) Kim et al. (2015) Mansour et al. (2014) Mariscotti et al. (2014) Rafferty et al. (2014) Spangler et al. (2011) Tagliafico et al. (2017) Taskin et al. (2017)	Amer et al. (2017) Ariaratnam et al. (2018) Bansal & Young (2015) Bian et al. (2016) Cai et al. (2015) Gennaro et al. (2013b) Kim et al. (2016) Luparia et al. (2013) Mariscotti et al. (2016) Mercier et al. (2015) Nakashima et al. (2017) Noroozian et al. (2011) Raghu et al. (2016) Refaat & Matar (2015) Schrading et al. (2014) Seo et al. (2014) Taha Ali et al. (2016) Takamoto et al. (2013) Tang et al. (2017) Yang et al. (2013) Zuley et al. (2013)	Alakhras et al. (2015) Bahrs et al. (2018) Berger et al. (2016) Chan et al. (2017) Choi et al. (2016b) Clauser et al. (2016) Destounis et al. (2013) Elizalde et al. (2016) Endo et al. (2018) Fornvik et al. (2018) Freer et al. (2015) Gennaro et al. (2013a) Helal et al. (2017) Kamal et al. (2016) Lång et al. (2014) Mariscotti et al. (2017) Munir et al. (2016) Tagliafico et al. (2015a) Tagliafico et al. (2015b) Thomassin-Naggara et al. (2015) Waldherr et al. (2015)

APPENDIX B: QUALITY ASSESSMENT FOR INCLUDED STUDIES

AMSTAR2 Tool for systematic reviews and meta-analysis

Phi et al., 2018

AMSTAR2 TOOL QUESTION		Answer	Comment
1	Did the research questions and inclusion criteria for the review include the components of the PICO?	Yes	Included studies set in both diagnostic and screening settings and which reported on at least one of four outcomes: CDR, recall rate, sensitivity, ad/or specificity); studies included at least 100 women with dense breasts
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?	Yes	Used PRISMA guidelines, discordance between reviewers and consensus reached or mediated by a third reviewer
3	Did the review authors explain their selection of the study designs for inclusion in the review?	Yes	
4	Did the review authors use a comprehensive literature search strategy?	Not sure	Only looked at PubMed and Scopus (Jan 2017 – May 2017) plus a manual bibliography check of included articles
5	Was there duplicate study selection and data extraction?	Not sure	
6	Did the review authors provide a list of excluded studies and justify the exclusion?	No	Exclusion criteria were described: did not contain original data, simulation studies
7	Did the review authors describe the included studies in adequate detail?	Yes	
8	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	Modified QUADAS-2 used by two reviewers independently: domains used were patient selection, index test, reference standard, flow and timing, applicability

AMSTAR2 TOOL QUESTION		Answer	Comment
9	Did the review authors report on the sources of funding for the studies included in the review?	No	
10	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	Yes	Random effects model (RevMan 5.3); analysis completed separately on screening and diagnostic populations; sub-group analysis completed to examine effect of covariates, modality, reading protocol and outcome; heterogeneity was quantified with I^2 for CDR and recall rate
11	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Not sure	RoB is not discussed
12	Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?	Not clear	RoB is not discussed
13	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes	
14	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?	NA	
15	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	No	

Yun et al., 2017

AMSTAR2 TOOL QUESTION		Answer	Comment
1	Did the research questions and inclusion criteria for the review include the components of the PICO?	Yes	Included studies in screening settings, which reported on
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?	Yes	Used MOOSE guidelines, disagreements between reviewers resolved by consensus
3	Did the review authors explain their selection of the study designs for inclusion in the review?	Yes	
4	Did the review authors use a comprehensive literature search strategy?	Yes	Studies published before 31 December 2016 using PubMed, EMBASE, and Cochrane Central and a manual bibliography check of included articles
5	Was there duplicate study selection and data extraction?	Yes	
6	Did the review authors provide a list of excluded studies and justify the exclusion?	No	Exclusion criteria were described: studies that intentionally controlled the number of participants from the eligible population, limited specific characteristics in the cohort, had sample sizes under 1000 in both arms, included data that overlapped with data published in other studies, or were interpreted without blinding to the results of the reference standard test.
7	Did the review authors describe the included studies in adequate detail?	Yes	Table of study characteristics and participant characteristics included
8	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	QUADAS-2 used by two reviewers independently: domains used were patient selection, index test, reference standard, flow and timing, and applicability.
9	Did the review authors report on the sources of funding for the studies included in the review?	No	

AMSTAR2 TOOL QUESTION		Answer	Comment
10	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	Yes	Random effects model (DerSimonian and Laird); analysis completed separately for overall cancer, invasive cancer, and carcinoma in situ; heterogeneity was quantified with I^2 , and subgroup or meta-regression analyses were applied to covariates suspected to be potential sources of heterogeneity
11	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Yes	
12	Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?	N	
13	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes	
14	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?	Not clear	
15	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Yes	Funding source and role of funding source provided: the researchers conducted the study independently of the funding sources. No conflicts of interest.

Garcia-Léon et al., 2015

AMSTAR2 TOOL QUESTION		Answer	Comment
1	Did the research questions and inclusion criteria for the review include the components of the PICO?	Yes	Included studies set in both diagnostic and screening settings and which reported on outcomes including sensitivity, specificity, predictive values odds ratio, Receiver Operating Characteristic, and intra- and inter-observer variability.
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?	Yes	Used PRISMA guidelines
3	Did the review authors explain their selection of the study designs for inclusion in the review?	Yes	
4	Did the review authors use a comprehensive literature search strategy?	Yes	Looked at MedLine, EMBASE, Web of Science, and PubMed (June 2010 to February 2013). Also included research from the Center for Reviews and Dissemination, the International Information Network on New and Emerging Health Technologies, the Cochrane Library. Websites reviewed included: the Spanish Ministry of Health, Social Services and Equality, the Platform of Agencies and Units of Evaluation of Health Technologies, the World Health Organisation, the Centers for Disease Control and Prevention, The Emergency Care and Research Institute, That National Institute for Health and Clinical Excellence and the American Cancer Society. A manual bibliography check of included articles was also performed.
5	Was there duplicate study selection and data extraction?	Yes	
6	Did the review authors provide a list of excluded studies and justify the exclusion?	No	Exclusion criteria included: descriptive-narrative reviews, letters to the editor, editorials, preclinical studies and preliminary studies with population contained in other studies.

AMSTAR2 TOOL QUESTION		Answer	Comment
7	Did the review authors describe the included studies in adequate detail?	Yes	Table of characteristics of study population included
8	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	Two reviewers independently assessed studies following the Cochrane Collaboration criteria and the QUADAS-2 tool: discrepancies were settled by consensus. Domains used were patient selection, index test, reference standard, follow-up, and applicability.
9	Did the review authors report on the sources of funding for the studies included in the review?	No	
10	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	NA	Narrative synthesis of results because it was not possible to achieve heterogeneity of the studies.
11	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	NA	
12	Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?	Not clear	
13	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes	
14	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?	NA	
15	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Yes	Authors stated funding sources and declared no conflicts of interest

Lei et al., 2014

AMSTAR2 TOOL QUESTION		Answer	Comment
1	Did the research questions and inclusion criteria for the review include the components of the PICO?	Yes	Included studies in diagnostic settings where patients had at least one breast lesion previously classified as BIRADS >2; more than 30 patients in the study; data that could be used to calculate sensitivity and specificity outcomes
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?	Not clear	
3	Did the review authors explain their selection of the study designs for inclusion in the review?	Yes	
4	Did the review authors use a comprehensive literature search strategy?	Yes	Looked at PubMed, EMBASE, the Cochrane Library, Web of Science, Chinese Biomedical Literature Database, Chinese scientific and technical journals, China Academic Journal database and Wanfang database (published between 1950 and June 2013) plus a manual bibliography check of included articles. MeSH terms were included
5	Was there duplicate study selection and data extraction?	Yes	
6	Did the review authors provide a list of excluded studies and justify the exclusion?	No	Exclusion criteria were described: review articles, letters, comments, editorials, conference abstracts and case reports
7	Did the review authors describe the included studies in adequate detail?	Yes	Table of study characteristics included.
8	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	QUADAS-2 used by two reviewers independently across 11 domains
9	Did the review authors report on the sources of funding for the studies included in the review?	No	

AMSTAR2 TOOL QUESTION		Answer	Comment
10	If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	Yes	Random effects model used; heterogeneity quantified using chi-squared value test and the inconsistency index of the diagnostic odds ratio value
11	If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Yes	
12	Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?	Yes	
13	Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes	
14	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?	No	Seven studies were included in the review and the authors cited the Cochrane handbook, which states that funnel plots should only be used to assess the publication bias for more than nine studies.
15	Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	No	