European Commission Initiative on Breast Cancer (ECIBC):
European guidelines on breast cancer screening and diagnosis
QUESTION

Should digital breast tomosynthesis vs. diagnostic mammography projections be used in the assessment of recalled women of average risk of breast cancer due to suspicious lesions at mammography screening?

RECOMMENDATION

The ECIBC's Guidelines Development Group suggests using digital breast tomosynthesis (DBT) over diagnostic mammography projections in the assessment of recalled women of average risk of breast cancer due to suspicious lesions at mammography screening (conditional recommendation, moderate certainty in the test accuracy data).
ASSESSMENT

POPULATION
Recalled women due to suspicious lesions at mammography screening

SETTING
Women of average risk of breast cancer with lesions suspicious for malignancy at mammography screening who are recalled for further imaging assessment

PERSPECTIVE
European Union

BACKGROUND
Digital breast tomosynthesis (DBT) is a modified mammographic technique which acquires low-dose projection images of the breast through a range of angles. This overcomes the effect of overlapping breast tissue, which is one of the limitations of full-field digital mammography (FFDM). (Michell MJ, 2012) When an underlying lesion is suspected in screening with FFDM, the woman is recalled for further assessment which among others could consist, depending on the setting, of spot compression, magnification, cleopatra view, cleavage view, mediolateral view. Recall leads to high costs and causes anxiety, not only prior to the assessment but also in the period to the subsequent screening mammogram (despite having received a final negative result in the assessment). (Brett J, 2001) Overlapping of normal tissues in FFDM may produce features on mammography which are suspicious for cancer, leading to recall for further tests. DBT could avoid this superimposition of normal tissue (Michell MJ, 2018) and therefore give the radiologist more certainty about the type and characteristics of the suspicious lesion. Based on the type of suspicious breast lesions, the following sub-populations were identified: architectural distortions, masses, asymmetric density, calcifications. Image findings were considered negative when BI-RADS 1-2, positive BI-RADS 3-4-5. BI-RADS 3 is recommended to be avoided in screening (D’Orsi CJ, 2013). BI-RADS 0 which itself indicates further assessment, was also considered as positive. Reference test was histology (surgery or biopsy) or long term follow-up.

Management of Conflicts of Interests (CoI): CoIs of all Guidelines Development Group (GDG) members were assessed and managed by the Joint Research Centre (JRC) following an established procedure in line with European Commission rules. GDG member participation in the development of the recommendations was restricted, according to CoI disclosure. Consequently, for this particular question, the following GDG members were recused from voting: Axel Gräwingholt and Elsa Pérez Gómez for declared interests. Miranda Langendam, as external expert, was also not allowed to vote, according to the ECIBC rules of procedure. For more information please visit: http://ecibc.jrc.ec.europa.eu/gdg-documents

http://healthcare-quality.jrc.ec.europa.eu
Is the problem a priority?

- No
- Probably no
- ❌ Yes
- Varies
- Don’t know

Digital breast tomosynthesis (DBT) is a modified mammographic technique which acquires low-dose projection images of the breast through a range of angles. This overcomes the effect of overlapping breast tissue, which is one of the limitations of full-field digital mammography (FFDM). (Michell MJ, 2012)

When an underlying lesion is suspected in screening with FFDM, the woman is recalled for further assessment which among others could consist, depending on the setting, of spot compression, magnification, cleopatra view, cleavage view, mediolateral view.

Recall leads to high costs and causes anxiety, not only prior to the assessment but also in the period to the subsequent screening mammogram (despite having received a final negative result in the assessment) (Brett J, 2001)

Overlapping of normal tissues in FFDM may produce features on mammography which are suspicious for cancer, leading to recall for further tests. DBT could avoid this superimposition of normal tissue (Michell MJ, 2018) and therefore give the radiologist more certainty about the type and characteristics of the suspicious lesion.

Additional considerations

The GDG prioritised this question for the ECIBC.
### How accurate is the test?

- **Very inaccurate**
- **Inaccurate**
- **Accurate**
- **Very accurate**
- **Varies**
- **Don’t know**

#### All consecutive studies

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 1000 patients tested (95% CI)</th>
<th>? of participants (studies)</th>
<th>Certainty of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prevalence 21%</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digital Breast Tomosynthesis</td>
<td>Assessment mammography</td>
<td>Digital Breast Tomosynthesis</td>
<td>Assessment mammography</td>
</tr>
<tr>
<td>True positives patients with breast cancer</td>
<td>202 (187 to 208)</td>
<td>193 (181 to 202)</td>
<td>86 (80 to 89)</td>
</tr>
<tr>
<td>False negatives patients incorrectly classified as not having breast cancer</td>
<td>8 (2 to 23)</td>
<td>17 (8 to 29)</td>
<td>4 (1 to 10)</td>
</tr>
<tr>
<td>True negatives patients without breast cancer</td>
<td>600 (498 to 672)</td>
<td>553 (450 to 640)</td>
<td>692 (573 to 774)</td>
</tr>
<tr>
<td>False positives patients incorrectly classified as having breast cancer</td>
<td>190 (118 to 292)</td>
<td>237 (150 to 340)</td>
<td>218 (136 to 337)</td>
</tr>
</tbody>
</table>

http://healthcare-quality.jrc.ec.europa.eu


1. There was no evidence of publication bias in the diagnostic forest plot, nor in the Decks test (p value 0.34).

2. The absolute differences are the additional cases identified or missed with digital breast tomosynthesis compared to additional mammographic views among those women recalled at the screening mammography assessment.

3. One study (Gilbert 2015), performed a retrospective analysis comparing 2DM versus DBT plus synthesised two view mammography. The remaining studies compared DBT versus diagnostic two view mammography.

4. Observed heterogeneity is explained by the use of different thresholds to define positive results (i.e. different classification systems and cut-off points) and to a lesser extent to the use of diverse additional imaging tests or reference standards.

5. In some of the included studies there was a non-blinded reading of the index tests. There was variability in how the evaluations were performed, in some cases they included additional tests such as ultrasound or special mammographic views. Those additional exams might be requested at clinical discretion and therefore could be a source of differential misclassification in the tests accuracy estimates.

6. There is large heterogeneity in the specificity estimates across studies.
Subgroup: non calcified lesions

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 1000 patients tested (95% CI)</th>
<th>? of participants (studies)</th>
<th>Certainty of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prevalence 18%</td>
<td></td>
<td>MODERA TE 1,2,3,4,5,a,b,c,d</td>
</tr>
<tr>
<td></td>
<td>Prevalence 8%</td>
<td></td>
<td>MODERA TE 1,2,3,4,5,a,b,c,d,e</td>
</tr>
<tr>
<td>Digital Breast Tomosynthesis</td>
<td>Assessment Mammography</td>
<td>Digital Breast Tomosynthesis</td>
<td>Assessment Mammography</td>
</tr>
<tr>
<td>True positives patients with breast cancer</td>
<td>169 (155 to 175)</td>
<td>164 (151 to 171)</td>
<td>75 (69 to 78)</td>
</tr>
<tr>
<td>False negatives patients incorrectly classified as not having breast cancer</td>
<td>11 (5 to 25)</td>
<td>16 (9 to 29)</td>
<td>5 (2 to 11)</td>
</tr>
<tr>
<td>True negatives patients without breast cancer</td>
<td>681 (517 to 763)</td>
<td>631 (492 to 730)</td>
<td>764 (580 to 856)</td>
</tr>
<tr>
<td>False positives patients incorrectly classified as having breast cancer</td>
<td>139 (57 to 303)</td>
<td>189 (90 to 328)</td>
<td>156 (64 to 340)</td>
</tr>
</tbody>
</table>

1. In some of the included studies there was a non-blinded reading of the index tests. There was variability in how the evaluations were performed, in some cases they included additional tests such as ultrasound or special mammographic views. Those additional exams might be requested at clinical discretion and therefore could be a source of differential misclassification in the tests accuracy estimates.
2. There was no evidence of publication bias in the diagnostic forest plot, nor in the Deeks test (p value 0.34).
3. One study (Gilbert 2015), performed a retrospective analysis comparing 2DM versus DBT plus synthetized two view mammography. The remaining studies compared DBT versus diagnostic two view mammography.
4. Observed heterogeneity is explained by the use of different thresholds to define positive results (i.e. different classification system and cut-off points) and to a lesser extent
to the use of adverse additional imaging tests or reference standards.

5. There is large heterogeneity in the specificity estimates across studies.


http://healthcare-quality.jrc.ec.europa.eu
Subgroup: calcified lesion

<table>
<thead>
<tr>
<th>Test result</th>
<th>Number of results per 1000 patients tested (95% CI)</th>
<th>% of participants (studies)</th>
<th>Certainty of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence 27%</td>
<td>Digital Breast Tomosynthesis       Assessment Mammography       Digital Breast Tomosynthesis  Assessment Mammography</td>
<td>1027 (1)</td>
<td>MODERATE&lt;sup&gt;1,a,b,c,d&lt;/sup&gt;</td>
</tr>
<tr>
<td>True positives patients with breast cancer</td>
<td>238 (227 to 248)                  238 (227 to 248)                114 (109 to 120)                114 (109 to 120)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 more TP in Digital Breast Tomosynthesis</td>
<td>0 more TP in Digital Breast Tomosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False negatives patients incorrectly classified as not having breast cancer</td>
<td>32 (22 to 43)                        32 (22 to 43)                16 (10 to 21)                   16 (10 to 21)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 fewer FN in Digital Breast Tomosynthesis</td>
<td>0 fewer FN in Digital Breast Tomosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>True negatives patients without breast cancer</td>
<td>285 (263 to 314)                  226 (204 to 248)                339 (313 to 374)                270 (244 to 296)</td>
<td>1027 (1)</td>
<td>MODERATE&lt;sup&gt;1,a,b,c,d&lt;/sup&gt;</td>
</tr>
<tr>
<td>59 more TN in Digital Breast Tomosynthesis</td>
<td>69 more TN in Digital Breast Tomosynthesis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>False positives patients incorrectly classified as having breast cancer</td>
<td>445 (416 to 467)                  504 (482 to 526)                531 (496 to 557)                600 (574 to 626)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59 fewer FP in Digital Breast Tomosynthesis</td>
<td>69 fewer FP in Digital Breast Tomosynthesis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. The study (Gilbert 2015), performed a retrospective analysis comparing 2DM versus DBT plus synthesized two view mammography.

2. Publication bias was not formally assessed as only one study was included.

3. Retrospective study from imaging records. Accuracy was estimated based on “decision to recall” from each reader which can not be considered a standard threshold. Reference standards were determined over the prospective original assessment of patients which did not include clinical follow-up.

4. The absolute differences are the additional cases identified or missed with digital breast tomosynthesis compared to additional mammographic views among those women recalled at the screening mammography assessment.
Additional considerations

The GDG agreed to focus on the 21% prevalence (estimated based on the included studies), more similar to the European setting.

The GDG agreed that for calcified and non-calcified lesions the results are also accurate, although the GDG focused on the overall results rather than the subgroup analysis.
How substantial are the desirable anticipated effects?

- Trivial
- Small
- Moderate
- Large
- Varies
- Don't know

Additional considerations

The GDG agreed that the desirable effects are moderate.
How substantial are the undesirable anticipated effects?

- Large
- Moderate
- Small
- Trivial
- Varies
- Don't know

**Additional considerations**

Radiation dose may be a concern. The comparison with supplementary views was not very clear. The radiation dose was explored in screening, but when carrying out additional projections on women who have been recalled, it is not clear how many views are done because it is very dependent on the patient.

As agreement was not reached by consensus with the GDG, voting was conducted among members without CoI with the following results:

11 members voted "trivial" and 6 members voted "small".
**What is the overall certainty of the evidence of test accuracy?**

- Very low
- Low
- **Moderate**
- High
- No included studies

*Additional considerations*

According to the table above, the GDG agreed the certainty of the evidence of test accuracy was moderate.
What is the overall certainty of the evidence for any critical or important direct benefits, adverse effects or burden of the test?

- Very low
- Low
- Moderate
- High
- No included studies

**Additional considerations**

The GDG considered what the consequences for clinical outcomes might be, based on the results with respect to test accuracy, as little information was presented on clinical outcomes. Only one study reported radiation dose, and expert knowledge on the GDG indicated that side effects were likely to be trivial. It was noted that there are some other studies which were not reviewed here, which would be consistent with the GDG’s considerations.
What is the overall certainty of the evidence of effects of the management that is guided by the test results?

〇 Very low
〇 Low
〇 Moderate
〇 High
☒ No included studies

Additional considerations

In this case there is again no evidence and it is an extrapolation, but the GDG is confident that the follow-up of women in terms of their management, given the test accuracy results, is leading to more benefits. Women who are correctly identified as true positives and true negatives are easier to manage.
How certain is the link between test results and management decisions?

- Very low
- Low
- Moderate
- High
- No included studies

Additional considerations

There are no included studies, but GDG is confident that there is proper action and management following from the tests applied. That is, as this is an intervention following assessment, if a biopsy needs to be done after the imaging, the GDG is certain that it will be done.
**What is the overall certainty of the evidence of effects of the test?**

- Very low
- Low
- Moderate
- High
- No included studies

**Additional considerations**

Although there are no included studies, the GDG agrees that there is moderate certainty for test accuracy and they are confident that the actions falling from the test accuracy data are appropriate even if no other evidence was evaluated.
<table>
<thead>
<tr>
<th><strong>Is there important uncertainty about or variability in how much people value the main outcomes?</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>○ Important uncertainty or variability</td>
</tr>
<tr>
<td>○ Possibly important uncertainty or variability</td>
</tr>
<tr>
<td>☒ Probably no important uncertainty or variability</td>
</tr>
<tr>
<td>○ No important uncertainty or variability</td>
</tr>
<tr>
<td>○ No known undesirable outcomes</td>
</tr>
</tbody>
</table>

**Additional considerations**

The GDG agreed that there was probably no important uncertainty on how much people value the main outcomes.
**Does the balance between desirable and undesirable effects favor the intervention or the comparison?**

- [ ] Favors the comparison
- [ ] Probably favors the comparison
- [ ] Does not favor either the intervention or the comparison
- [x] Probably favors the intervention
- [ ] Favors the intervention
- [ ] Varies
- [ ] Don't know

**Additional considerations**

The GDG agreed that the balance of effects probably favours the intervention.
**How large are the resource requirements (costs)?**

<table>
<thead>
<tr>
<th>Option</th>
<th>Additional considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large costs</td>
<td></td>
</tr>
<tr>
<td>Moderate costs</td>
<td>There was no research evidence but the GDG agreed on the following.</td>
</tr>
<tr>
<td>Negligible costs and savings</td>
<td></td>
</tr>
<tr>
<td>Moderate savings</td>
<td></td>
</tr>
<tr>
<td>Large savings</td>
<td></td>
</tr>
<tr>
<td>Varies</td>
<td></td>
</tr>
<tr>
<td>Don't know</td>
<td></td>
</tr>
</tbody>
</table>

**With regards to resources:**

- The DBT-device is much more expensive than the magnification-device.

- The time needed by the radiographer is similar: for magnification the radiographer needs more time for positioning, but on the other hand, for DBT it takes longer to take the image

- For the radiologist, the interpretation time of DBT may be longer than that needed for magnification.
With regards to unitary costs:

- DBT cost (one/two views, one/two sides) varies from 50 to 190 Euros in Germany and from 24 to 130 Euros in Spain (data provided by GDG members, 2018).

- Magnification is cheaper, additional images, one view or more costs approximately 10 Euros in Germany and Spain (data provided by GDG members, 2018).

The GDG agreed the resources required for DBT would be moderate.
What is the certainty of the evidence of resource requirements (costs)?

- Very low
- Low
- Moderate
- High
- No included studies

Additional considerations

Raw estimates for base costs were looked for by GDG members, so the GDG agreed the certainty of the evidence would be very low.
Does the cost-effectiveness of the intervention favor the intervention or the comparison?

- Favors the comparison
- Probably favors the comparison
- Does not favor either the intervention or the comparison
- Probably favors the intervention
- Favors the intervention
- Varies
- No included studies

Additional considerations

No included studies.

No studies were included.
**What would be the impact on health equity?**

- Reduced
- Probably reduced
- Probably no impact
- Probably increased
- Increased
- Varies
- Don't know

**Additional considerations**

The GDG agreed that equity would vary. There may be differences within the target countries with regard to availability of the tomosynthesis equipment and whether or not it is included in the screening programmes.
<table>
<thead>
<tr>
<th>Is the intervention acceptable to key stakeholders?</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ] No</td>
</tr>
<tr>
<td>[ ] Probably no</td>
</tr>
<tr>
<td>[x] Probably yes</td>
</tr>
<tr>
<td>[ ] Yes</td>
</tr>
<tr>
<td>[ ] Varies</td>
</tr>
<tr>
<td>[ ] Don't know</td>
</tr>
</tbody>
</table>

**Additional considerations**

The GDG agreed that acceptability would probably vary as, for example for those who pay, cost may create barriers.
Is the intervention feasible to implement?

- No
- Probably no
- Probably yes
- Yes
- Varies
- Don’t know

Additional considerations

The GDG agreed the feasibility would vary depending on the country, but the GDG agreed that this variability would be reduced as time passes. Currently it is difficult to acquire a new mammography device without the possibility of implementing tomosynthesis. It is just a question of time and turnover.
## CONCLUSIONS

Should digital breast tomosynthesis vs. diagnostic mammography projections be used in the assessment of recalled women of average risk of breast cancer due to suspicious lesions at mammography screening?

<table>
<thead>
<tr>
<th>TYPE OF RECOMMENDATION</th>
<th>Strong recommendation against the intervention</th>
<th>Conditional recommendation against the intervention</th>
<th>Conditional recommendation for either the intervention or the comparison</th>
<th>Conditional recommendation for the intervention</th>
<th>Strong recommendation for the intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>🍁</td>
<td>🍁</td>
<td>🍁</td>
<td>🍁</td>
<td>🍁</td>
</tr>
</tbody>
</table>

### RECOMMENDATION

The ECIBC’s Guidelines Development Group suggests using digital breast tomosynthesis (DBT) over diagnostic mammography projections in the assessment of recalled women of average risk of breast cancer due to suspicious lesions at mammography screening (conditional recommendation, moderate certainty in the test accuracy data).

### JUSTIFICATION

Overall justification

The GDG agreed the recommendation by consensus.

It is a conditional recommendation mainly due to the availability of the devices. In addition, despite moderate certainty in the high accuracy of the test results, and a balance that probably favours DBT, there are concerns about the associated moderate costs, increased training needed to correctly use this technology, absence of cost-effectiveness data and feasibility of its implementation.
The GDG agreed that this recommendation applies to both subgroups of patients examined, those with calcified lesions and those with non calcified lesions.

None were considered by GDG.

Quality control procedures and quality standards should be further developed. Standards should be developed in particular for the image quality of synthesised 2D images from the tomosynthesis technology.

Ultrasound is often included in the management of assessment after a positive finding in screening mammography. Further research should be conducted exploring which subgroups would avoid ultrasound after DBT-additional projections, as well as which lesions (usually masses) are assessed with ultrasound instead of additional projections/DBT. The use of DBT in high mammographic breast density should be explored, that is, whether or not accuracy results are affected by breast density. Members of the GDG raised the question of whether the evidence discussed applies to repeated assessments too, but there was no agreement in the GDG if this is a research priority. Whether or not to use one or two views for tomosynthesis in assessment should be explored.