ORIGINAL ARTICLE

Digital tomosynthesis in breast cancer: A systematic review

F.J. García-León*, A. Llanos-Méndez, R. Isabel-Gómez

Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, Consejería de Igualdad, Salud y Políticas Sociales, Sevilla, Spain

Received 21 March 2014; accepted 22 June 2014

KEYWORDS
Breast cancer; Screening; Tomosynthesis; Three-dimensional imaging; Mammography; Sensitivity and specificity

Abstract
Objective: To estimate and compare the diagnostic validity of tomosynthesis and digital mammography for screening and diagnosing breast cancer.

Material and methods: We systematically searched MedLine, EMBASE, and Web of Science for the terms breast cancer, screening, tomosynthesis, mammography, sensitivity, and specificity in publications in the period comprising June 2010 through February 2013. We included studies on diagnostic tests and systematic reviews. Two reviewers selected and evaluated the articles. We used QUADAS 2 to evaluate the risk of bias and the NICE criteria to determine the level of evidence. We compiled a narrative synthesis.

Results: Of the 151 original studies identified, we selected 11 that included a total of 2475 women. The overall quality was low, with a risk of bias and follow-up and limitations regarding the applicability of the results. The level of evidence was not greater than level II. The sensitivity of tomosynthesis ranged from 69% to 100% and the specificity ranged from 54% to 100%. The negative likelihood ratio was good, and this makes tomosynthesis useful as a test to confirm a diagnosis. One-view tomosynthesis was no better than two-view digital mammography, and the evidence for the superiority of two-view tomosynthesis was inconclusive.

Conclusions: The results for the diagnostic validity of tomosynthesis in the diagnosis of breast cancer were inconclusive and there were no results for its use in screening.

© 2014 SERAM. Published by Elsevier España, S.L.U. All rights reserved.

PALABRAS CLAVE
Cáncer de mama; Cribado; Tomosíntesis;

Tomosíntesis digital en el cáncer de mama. Revisión sistemática

Resumen
Objetivo: Estimar y comparar la validez diagnóstica de la tomosíntesis y la mamografía digital para cribar y diagnosticar el cáncer de mama.

* Corresponding author.
E-mail address: fjavier.garcia.leon@juntadeandalucia.es (F.J. García-León).
Material y métodos: Realizamos una revisión sistemática consultando MedLine, EMBASE y Web of Science en el periodo de junio de 2010 a febrero de 2013. Los términos de búsqueda fueron: cáncer de mama, cribado, tomosíntesis, mamografía, sensibilidad y especificidad. Se incluyeron estudios de pruebas diagnósticas y revisiones sistemáticas. Dos investigadores hicieron la selección y evaluación. Usamos QUADAS 2 para valorar el riesgo de sesgo y los criterios NICE para el nivel de evidencia. Se hizo una síntesis narrativa.

Resultados: De los 151 estudios originales identificados se seleccionaron 11 que incluyeron 2.475 mujeres. Su calidad fue baja, con riesgo de sesgo de selección y seguimiento, y limitaciones para aplicar sus resultados. Su nivel de evidencia no fue superior a II. La sensibilidad de la tomosíntesis osciló entre el 69 y el 100% y la especificidad entre el 54 y el 100%. El cociente de probabilidad negativo fue bueno, lo que la convertiría en una prueba de confirmación diagnóstica. La tomosíntesis con una proyección no fue superior a la mamografía digital con 2, y con 2 proyecciones los resultados no fueron concluyentes.

Conclusiones: Los resultados de la validez diagnóstica de la tomosíntesis en el diagnóstico del cáncer de mama no fueron concluyentes, y no los hubo para usarla en el cribado.

© 2014 SERAM. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

Introduction

Mammogram is the basic modality for breast cancer clinical diagnosis and screening. Screening continues to be the main preventive measure to reduce mortality though its effect is subject to discussion. In Spain, it is recommended for 50–69-year-old women, though some autonomous communities also include 45–49 year-old women. In spite of its high sensitivity (between 85 and 95%) and specificity (>90%) mammograms also give false negatives and positives that bring about anxiety, unnecessary procedures and overdiagnoses. To this we have to add the effects of radiation and the discomfort caused by breast compression. Diagnostic limitations are greater in dense breasts which are in turn the ones with the highest risk for developing cancer.

Digital mammographies have operational advantages and possibilities for technological evolution. With them, cancer detection rate is slightly higher than that of conventional mammographs yet recall rates or the characteristics of the tumors found do not usually vary. Digital breast tomosynthesis has developed from digital mammographs as an alternative or complement. It was approved by European Commission (EC) in 2008 and by the U.S. Food and Drug Administration (FDA) in 2011 and it is installed in 12 health care centers in Spain. It differs from digital mammographies in that for each projection the X-ray tube describes a rotation arch on a plane around the breast ranging from 11 to 50°, taking between 9 and 25 images. The images are computed processed for a digital breast reconstruction in 3 dimensions. There is a great variety of machines for tomosynthesis. With a single breast compression some machines successively obtain two-dimensional digital mammographic images and three-dimensional tomosynthesis images, while others obtain the 3D images directly from tomosynthesis from which they reconstruct a 2D image (synthesized image). Applications for the analysis of texture, to make numerical quantifications and detect and help to diagnose masses and microcalcifications have been developed to analyze the images.

A systematic review was conducted in the year 2010 to evaluate the diagnostic validity of tomosynthesis, which demonstrated its possible utility for the diagnosis of breast cancer but without evidence about its utility in screening. Due to the increase of medical literature on this regard and the growing interest among professionals we thought it was a good idea to repeat it in order to update the evidence available to establish its effectiveness, in terms of diagnostic validity and accuracy, in screening and breast cancer diagnosis.

Material and methods

This systematic literature review was conducted following the PRISMA statement recommendations and devising an internal work protocol. The results were synthesized in a narrative manner because it was not possible to achieve statistic combination due to the heterogeneity of the studies.

Sources of information

The MedLine, EMBASE, Web of Science and PubMed (Annex 1) databases were reviewed (from June 2010 to February 2013). Research was also conducted at the Center for Reviews and Dissemination (CRD), the International Information Network on New and Emerging Health Technologies (EuroScan) and the Cochrane Library. The websites of agencies not included in INAHTA were reviewed, the Spanish Ministry of Health, Social Services and Equality, the platform of Agencies and Units of Evaluation of Health Technologies (AUnETS), the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), The Emergency Care Research Institute (ECRI), The National Institute for Health and Clinical Excellence (NICE) and the American Cancer Society (http://www.cancer.org). Also a manual review of the bibliography of the studies included was implemented.
Both natural and controlled languages were used to consult the databases; independent strategies were built for each resource adapting the consultations to the controlled language typical of each base. Strategies were based on the following terms: breast neoplasms, screening, tomosynthesis, mammogram, sensitivity and specificity, among others.

**Selection of studies**

Two trained skilled researchers reviewed both the titles and abstracts of the publications to select the ones that included women with clinical suspicion of breast cancer or who were included in the disease’s screening programs, whose results of validity and diagnostic accuracy were compared to digital mammographies using anatomicopathological studies or follow-up as reference standards. Descriptive-narrative reviews were precluded as well as letters to the editor, editorials, preclinical studies and preliminary studies with population was contained in other studies that would not provide with relevant results or incomplete studies where it would not be possible to assess quality adequately.

**Data mining**

Information was collected by one reviewer only. The data collection included variables related with the characteristics of the studies (author, country, publication year, objectives, stage of the study according to Sackett’s scale), the population (number, age and indication of test, as well as inclusion or not of calcifications), technique used to obtain the images (number of projections, arch described, radiation dose, time used and size of the tomogram) as well as the brand and model, workstation for the reading and the number and training of the corresponding radiologist. The results included diagnostic validity and precision parameters (sensitivity, specificity, predictive values, odds ratios, ROC curve and intra- and inter-observer variability). Whenever possible, 2×2 contingency tables were built to calculate these parameters from the data in each study (Table 1).

**Bias risk assessment and level of evidence**

Two reviewers assessed biases independently by establishing the quality of the original studies following the Cochrane Collaboration criteria and the QUADAS-2 tool. Discrepancies were settled by consensus. The method used to establish the level of evidence was based on the NICE criteria.

**Results**

**Results after reviewing the medical literature**

One hundred and fifty-one bibliographic references were collected. After the selection process based on the inclusion and exclusion criteria, 11 full-text studies were analyzed (Fig. 1) with a prospective design corresponding to Stage III (cohorts) according to Sackett’s
criteria, except one in Stage II (case–control).\textsuperscript{19} Works were conducted in Sweden, Italy, USA, Switzerland, Norway and the United Kingdom.

Description of the population
2475 women were included, 1094 with an mean age of 51 years with radiological alterations in the screening,\textsuperscript{15-16} 356 with an mean age of 57–60 years with clinical signs\textsuperscript{19,20,23} and 1025 with an mean age of 51–57 years in some of the aforementioned situations without specification.\textsuperscript{17,18,21,22} Women with calcifications were precluded in 3 studies\textsuperscript{15,16,21} while another study included women with dense breasts only.\textsuperscript{22}

Description of the equipment/machines used
6 Selenia Dimensions\textsuperscript{8} machines and 5 prototypes were used (Table 2) and they were able to obtain single-projection tomosynthesis images (oblique mid lateral),\textsuperscript{17,19,20} 2 projections (oblique mid lateral and cranial-caudal),\textsuperscript{14-16,21} and indistinctly 1 or 2\textsuperscript{21,22}; in 2 studies the number was not specified.\textsuperscript{13,18} The machines obtained 11 images for each projection,\textsuperscript{14,17,21} or between 15 and 25\textsuperscript{15,19,20,22,23}; the information was not provided in 3 of them.\textsuperscript{13,15,18} The x-ray tube described a 15° arch in most of them though extreme values ranged from 11 to 60°. The image was reconstructed through 1 mm-thick tomograms in one study at 3 mm,\textsuperscript{22} and it was not described in other 2 studies.\textsuperscript{13,19,21} The number of radiologists who interpreted the results in each study ranged from 2 to 27, and their experience in breast radiological diagnosis ranged between 1 and 30 years. The radiologists of two studies had experience in tomosynthesis\textsuperscript{15,23} and the others had been trained in short seminar or studying between 25 and 150 cases.

Reference test in the studies
All the studies used a double reference standard (biopsy or follow-up). The patients with positive radiological findings were studied anatomopathologically and those who tested negative were followed up for 12–16 months (range 6–39).

Outcome measurements used in the studies
The outcomes included sensitivity, specificity, predictive values, odds ratios and area under the curve (AUC) Receiver Operating Characteristic (ROC) (Table 3). Also we compared the results of tomosynthesis with those of digital mammographies,\textsuperscript{13,14,17,21,22} including on occasions compression modalities,\textsuperscript{15,16,18,23} magnification\textsuperscript{18} and ultrasound.\textsuperscript{18-20}

Bias risk in the studies
The quality of the studies was generally poor showing a high risk of bias due to problems in the selection, which in turn
<table>
<thead>
<tr>
<th>Author</th>
<th>Machine</th>
<th>Projections</th>
<th>Arch</th>
<th>Dose</th>
<th>Time</th>
<th>Tomograms</th>
<th>Projections</th>
<th>Working station</th>
<th>Radiologists (training in tomosynthesis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernardi et al., 2011¹³</td>
<td>Selenia dimensions. COMBO. Hologic</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>7 (short seminar)</td>
</tr>
<tr>
<td>Michell et al., 2012¹⁴</td>
<td>Selenia dimensions. Hologic</td>
<td>11</td>
<td>15°</td>
<td>Low</td>
<td>ND</td>
<td>1 mm</td>
<td>Two</td>
<td>ND</td>
<td>5 (ND)</td>
</tr>
<tr>
<td>Tagliafico et al., ¹⁵</td>
<td>Selenia Dimensions. Hologic</td>
<td>ND</td>
<td>15°</td>
<td>Ultra Low</td>
<td>ND</td>
<td>1 mm</td>
<td>Two</td>
<td>ND</td>
<td>2 (experimented)</td>
</tr>
<tr>
<td>Brandt et al., 2013¹⁶</td>
<td>Selenia dimensions. Hologic*</td>
<td>15</td>
<td>ND</td>
<td>Low</td>
<td>ND</td>
<td>1 mm</td>
<td>Two</td>
<td>SecurView 6. Hologic Barco Coronis 5 MP Mammo Hologic</td>
<td>3 (8 hours)</td>
</tr>
<tr>
<td>Waldherret et al., 2013¹⁷</td>
<td>Selenia dimensions. Hologic</td>
<td>11</td>
<td>15°</td>
<td>Same as the digital mammogram</td>
<td>4 s</td>
<td>1 mm</td>
<td>One</td>
<td>Sun Microsystem Ultra 24 Workstation GE SenoAdvantage</td>
<td>2 (ND)</td>
</tr>
<tr>
<td>Skaane et al., 2012¹⁸</td>
<td>Selenia dimensions. Hologic</td>
<td>ND</td>
<td>15°</td>
<td>Low</td>
<td>ND</td>
<td>1 mm</td>
<td>ND</td>
<td>Sun Microsystem Ultra 24 Workstation GE SenoAdvantage</td>
<td>4 (short seminar)</td>
</tr>
<tr>
<td>Svahnet et al., 2012¹⁹</td>
<td>Prototype Mammat NovationDR. Siemens</td>
<td>25</td>
<td>20°</td>
<td>Same or lower than the digital mammogram</td>
<td>20 s</td>
<td>ND</td>
<td>One</td>
<td>Sun Microsystem Ultra 24 Workstation GE SenoAdvantage</td>
<td>5 (30 cases)</td>
</tr>
<tr>
<td>Gennaro et al., 2010²⁰</td>
<td>Prototype Senographe DS. General electric</td>
<td>15</td>
<td>20°</td>
<td>Same or lower than the digital mammogram</td>
<td>ND</td>
<td>1 mm</td>
<td>One</td>
<td>GE SenoAdvantage</td>
<td>6 (25 cases)</td>
</tr>
<tr>
<td>Rafferty et al., ²¹</td>
<td>Prototype Hologic</td>
<td>11</td>
<td>15°</td>
<td>Same as the digital mammogram</td>
<td>10 s</td>
<td>ND</td>
<td>Two</td>
<td>ND</td>
<td>27 (150 cases)</td>
</tr>
<tr>
<td>Wallis, 2012²²</td>
<td>Prototype Microdose mammogram. Sectra</td>
<td>21</td>
<td>11°</td>
<td>Same as the digital mammogram</td>
<td>ND</td>
<td>3 mm</td>
<td>One or two</td>
<td>ND</td>
<td>20 (ND)</td>
</tr>
<tr>
<td>Noroozian et al., 2012²³</td>
<td>Prototype tomosynthesis and ultrasounds general electric</td>
<td>21</td>
<td>60°</td>
<td>1.4 times higher than the mammogram</td>
<td>8 s</td>
<td>1 mm</td>
<td>One or two</td>
<td>IBM prototype T221</td>
<td>4 (experimented)</td>
</tr>
</tbody>
</table>

ND: not described; s: seconds. *: not in the U.S. market.
Table 3 Description of outcome measurements.

<table>
<thead>
<tr>
<th>Author</th>
<th>Level of evidence</th>
<th>Positive radiology</th>
<th>Months of follow-up</th>
<th>AUC ROC</th>
<th>Inter-observer differences</th>
<th>Rellamadas</th>
<th>Unidad de análisis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernardiet et al., 2011†</td>
<td>II</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Chi-square</td>
<td>Percentage</td>
<td>Women</td>
</tr>
<tr>
<td>Michellet et al., 2012†</td>
<td>II</td>
<td>RCRBG</td>
<td>Biopsies (18–36)</td>
<td>Hanley et al. method</td>
<td>ND</td>
<td>ND</td>
<td>Imagery</td>
</tr>
<tr>
<td>Tagliafico et al., 2012†</td>
<td>II</td>
<td>BI-RADS 3–5</td>
<td>Biopsies (12–16)</td>
<td>Criterion of non-inferiority</td>
<td>Kappa</td>
<td>ND</td>
<td>Women</td>
</tr>
<tr>
<td>Brandt et al., 2013†</td>
<td>II</td>
<td>BI-RADS 4 or 5</td>
<td>Biopsies (6–39)</td>
<td>ND</td>
<td>Kappa</td>
<td>ND</td>
<td>Imagery</td>
</tr>
<tr>
<td>Waldherr et al., 2013†</td>
<td>II</td>
<td>BI-RADS 4 or 5</td>
<td>Biopsies (12–16)</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Women</td>
</tr>
<tr>
<td>Skaaneet al., 2012†</td>
<td>III</td>
<td>ND</td>
<td>Biopsies (12)</td>
<td>Incomprehensible</td>
<td>ND</td>
<td>ND</td>
<td>Imagery</td>
</tr>
<tr>
<td>Gennaroet al., 2010†</td>
<td>II</td>
<td>BI-RADS 3–5</td>
<td>Biopsies (12)</td>
<td>Criterion of non-inferiority</td>
<td>ND</td>
<td>ND</td>
<td>Imagery</td>
</tr>
<tr>
<td>Raffertyet al., 2013†</td>
<td>II</td>
<td>ND</td>
<td>Biopsies (12)</td>
<td>ND</td>
<td>Difference</td>
<td>Women</td>
<td></td>
</tr>
<tr>
<td>Wallis, 2012†</td>
<td>II</td>
<td>ND</td>
<td>Biopsies (12)</td>
<td>Two-tailed P value</td>
<td>ND</td>
<td>Difference</td>
<td>Imagery</td>
</tr>
<tr>
<td>Noroozian et al., 2012†</td>
<td>II</td>
<td>BI-RADS 4 or 5</td>
<td>Biopsies (24)</td>
<td>Dorfman-Berbaum-Metz method</td>
<td>ND</td>
<td>ND</td>
<td>Imagery</td>
</tr>
</tbody>
</table>

AUC: area under the curve; ND: not described; RCRBG: Royal College of Radiologists Breast Group. *: level of evidence according to NICE. †: difference in the recall percentage between tomosynthesis and mammograms.

limited the applicability of the results (Fig. 2). The diagnostic validity of tomosynthesis can be overestimated because some studies include a high percentage of women with cancer, others include exclusively patients with high Breast Imaging Reporting and Data System (BI-RADS) and others preclude women with calcifications (in which tomosynthesis does not outperform mammogram). There were some limitations in the external validity due to the high percentage of women with cancer in the studies (between 10 and 80%), the unequal distribution of the disease and its risk factors in different countries, and the size of the sample used whose statistic power is analyzed in one of the studies only.21

The results of tomosynthesis were interpreted without knowing the reference standards; therefore the performance or interpretation bias was low. The machines were adequately described though there was great variability in terms of settings and characteristics of the workstation as in number, experience and training the radiologists which made it hard to compare and apply the results obtained.

The interpretation of the reference standard and the follow-up period were independent from the test being studied; this is why the incorporation bias was low. However all studies used a double reference standard based on the results of the test which entails a high risk of differential verification bias that would lead to overestimate the sensitivity and specificity of tomosynthesis.

There were important losses in half of the studies implying a high risk of wear whose significance we cannot assess since we do not know the specifics of such losses.

Ten studies showed level II of evidence and one study level III.

Study results

Diagnostic validity and test global performance

Sensitivity ranged between 69 and 100%, increasing as the prevalence of cancer dropped in the population in which 2 projections were taken; the variability of specificity was greater ranging between 54 and 100%. In most studies the positive odds ratio was acceptable and it was considered an excellent test in the larger sample study. It must be pointed out that most studies where the negative odds ratio was calculated had good or excellent values which gives the test a high capability to rule out the disease which in turn would make it a test for diagnostic confirmation (Table 4).

The AUC was calculated in 6 studies.14,15,19,20,22,23 The best results were obtained combining two-projection tomosynthesis plus a mammogram in patients in whom radiological alterations had been identified during the screenings with a value of 0.96 (0.95–0.97).14 The results worsened in patients with clinically diagnosed signs or symptoms who had one- or two-projection tomosynthesis: AUC 0.91.21 The lowest values were obtained when one projection was obtained only with AUC of 0.86 (0.80–0.91)19 and 0.85.22 In the study of women with dense breasts coming from screening and the clinical field the results were equally better both with 2 and 1-projection tomosynthesis (0.85 and 0.77 respectively).22
### Table 4 Results of diagnostic validity.

<table>
<thead>
<tr>
<th>Author</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
<th>POR (95% CI)</th>
<th>NOR (95% CI)</th>
<th>AUC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bernardiet et al., 2011</strong>&lt;sup&gt;13&lt;/sup&gt;</td>
<td>100(100–100)</td>
<td>74.4 (67.1–81.7)</td>
<td>37.5 (24.8–50.1)</td>
<td>100 (100–100)</td>
<td>3.9 (2.9–5.2)</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td><strong>Michell et al., 2012</strong>&lt;sup&gt;14&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M 5&lt;sup&gt;1&lt;/sup&gt;</td>
<td>57.8 (51.0–64.6)</td>
<td>99.6 (99.1–100)</td>
<td>98.3 (96.0–100)</td>
<td>86.5 (83.9–89.1)</td>
<td>86.5 (83.9–89.1)</td>
<td>160.5 (160.5–160.5)</td>
<td>0.42 (0.42–0.42)</td>
</tr>
<tr>
<td>M 4, 5&lt;sup&gt;1&lt;/sup&gt;</td>
<td>80.8 (75.4–86.2)</td>
<td>96.7 (95.2–98.2)</td>
<td>90.1 (85.8–94.4)</td>
<td>93.2 (91.1–95.2)</td>
<td>24.9 (15.7–39.4)</td>
<td>3.9 (2.9–5.2)</td>
<td>0.2 (0.2–0.2)</td>
</tr>
<tr>
<td>M 3, 4, 5&lt;sup&gt;1&lt;/sup&gt;</td>
<td>100 (100–100)</td>
<td>74.2 (70.6–77.8)</td>
<td>58.7 (53.6–63.9)</td>
<td>100 (100–100)</td>
<td>3.8 (3.3–4.4)</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td><strong>Tagliafico et al., 2012</strong>&lt;sup&gt;15&lt;/sup&gt;</td>
<td>100 (91–100)</td>
<td>100 (91–100)</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>1</td>
</tr>
<tr>
<td><strong>Brandt et al., 2013</strong>&lt;sup&gt;16&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs 1&lt;sup&gt;1&lt;/sup&gt;</td>
<td>100 (100–100)</td>
<td>93.5 (89.1–97.8)</td>
<td>50 (25.5–74.5)</td>
<td>100 (100–100)</td>
<td>15.3 (7.8–30.0)</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Obs 2&lt;sup&gt;1&lt;/sup&gt;</td>
<td>100 (100–100)</td>
<td>92.6 (88.08–97.29)</td>
<td>47.06 (23.33–70.79)</td>
<td>100 (100–100)</td>
<td>13.67 (7.29–25.63)</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Obs 3&lt;sup&gt;1&lt;/sup&gt;</td>
<td>87.50 (64.58–100.00)</td>
<td>89.43 (84.0–94.8)</td>
<td>35 (14.1–55.9)</td>
<td>99.1 (97.3–100.0)</td>
<td>8.2 (4.6–14.7)</td>
<td>0.1 (0.0–0.8)</td>
<td>ND</td>
</tr>
<tr>
<td><strong>Waldherr et al., 17</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total&lt;sup&gt;1&lt;/sup&gt;</td>
<td>88.3 (81.6–95.1)</td>
<td>79.3 (68.8–89.7)</td>
<td>86.3 (79.1–93.5)</td>
<td>82.1 (72.1–92.1)</td>
<td>82.1 (72.1–92.1)</td>
<td>4.2 (2.57–7.1)</td>
<td>0.1 (0.0–0.2)</td>
</tr>
<tr>
<td>Clinical</td>
<td>88.7</td>
<td>93.8</td>
<td>98.2</td>
<td>68.2</td>
<td>14.3</td>
<td>0.1</td>
<td>ND</td>
</tr>
<tr>
<td>Screening</td>
<td>87.5</td>
<td>73.2</td>
<td>65.6</td>
<td>90.9</td>
<td>3.26</td>
<td>0.2</td>
<td>ND</td>
</tr>
<tr>
<td>Low D</td>
<td>94.4</td>
<td>74.1</td>
<td>82.9</td>
<td>90.5</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>High D</td>
<td>84 (70.9–92.8)</td>
<td>83.9 (66.3–94.6)</td>
<td>89.4 (76.9–96.5)</td>
<td>76.5 (58.8–89.3)</td>
<td>76.5 (58.8–89.3)</td>
<td>4.6 (3.2–6.7)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Skaane et al., 2012</strong>&lt;sup&gt;18&lt;/sup&gt;</td>
<td>100 (100–100)</td>
<td>78.43 (70.4–86.4)</td>
<td>55.1 (41.1–69.0)</td>
<td>100 (100–100)</td>
<td>4.6 (3.2–6.7)</td>
<td>0</td>
<td>ND</td>
</tr>
<tr>
<td>Author</td>
<td>Sensitivity (95% CI)</td>
<td>Specificity (95% CI)</td>
<td>PPV (95% CI)</td>
<td>NPV (95% CI)</td>
<td>POR (95% CI)</td>
<td>NOR (95% CI)</td>
<td>AUC (95% CI)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
</tr>
<tr>
<td><em>Studies using prototypes</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Svahn et al., 2012(^{19})</td>
<td>89.7</td>
<td>54.4</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Gennaro et al., 2010(^{20})</td>
<td>69.8</td>
<td>88.9</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Rafferty et al., 2013(^{21})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 1</td>
<td>76.2</td>
<td>89.2</td>
<td>56.2</td>
<td>95.4</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Invasive 78.6</td>
<td>In situ 71.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study 2</td>
<td>78.7</td>
<td>84.5</td>
<td>50.1</td>
<td>95.3</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td></td>
<td>Invasive 82.3</td>
<td>In situ 70.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wallis, 2012(^{22})</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
<tr>
<td>Noroozian et al., 2012(^{23})</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td></td>
</tr>
</tbody>
</table>

AUC: area under the curve; NOR: negative odds ratio; POR: positive odds ratio; POR and NOR: between brackets CI for Taylor method; D: density; M: Royal College of Radiologists Breast Group Classification; ND: not described; Obs: observer; NPV: negative predictive value; PPV: positive predictive value.

\(^{1}\) Estimated by the authors.
Some results were not assessed because they came from a very small sample and others did not have any confidence intervals.\textsuperscript{15}

\textbf{Precision}

One of the works studied the variability of the tomosynthesis results among 7 radiologists\textsuperscript{13} and the percentage of true negatives was significantly different \((p = 0.004)\). Another study confirmed a high level concordance between the results of two experienced radiologists with a kappa index of 0.95.\textsuperscript{15}

\textbf{Comparison between tomosynthesis and digital mammograms}

In the 9 studies where the validity or precision of digital mammogram was compared to that of tomosynthesis\textsuperscript{14-17,19-23} the tomosynthesis showed better results when it was combined with the mammogram and when two projections were taken–not when only one was taken. The test’s global performance (measured as AUC) was significantly better with both the tomosynthesis and the mammogram combined than with the mammogram only.\textsuperscript{14-21} This was also the case for the two-projection tomosynthesis (both in masses and microcalcifications) but not with one.\textsuperscript{22} No differences between tomosynthesis and compression mammograms were seen in the works where microcalcifications were not precluded.\textsuperscript{15,16,23}

In one of the studies the sensitivity of tomosynthesis was greater but not for the fraction of false negatives (1-specificity).\textsuperscript{19} The results of the radiologists with tomosynthesis were better for the classification of masses or the location of lesions.\textsuperscript{17,19,23}

\textbf{Discussion}

During the last few years there has been a significant increase in the number of studies published about tomosynthesis yet its utility in the screening of the breast cancer population is still unknown. This situation is anticipated to change with the ongoing studies among them the Malmö Breast Tomosynthesis Screening Trial (MBTST) of 15,000 women comparing the mammogram and the tomosynthesis, and the Tomosynthesis in the Oslo Breast Cancer Screening Program (DBT) of 25,000 women. The preliminary results of the latter\textsuperscript{18} indicate that combining the mammogram and the two-projection tomosynthesis would increase cancer detection in the screenings and allow us to detect the most invasive types of cancer.\textsuperscript{18} One review that other than these results includes other outcomes about screening conducted in Italy has not found enough evidence for mammograms to be complemented with tomosynthesis.\textsuperscript{25} We need to be aware though that these trials like the UK TOMMY project and Michell et al.’s work\textsuperscript{14} do not imply replacing mammograms with tomosynthesis but use them together.

The most significant result, in our opinion, is that tomosynthesis has a good or excellent negative odds ratio, which makes it a useful test to rule out the disease and therefore a diagnostic confirmation test. Another important finding consistent with those of other studies\textsuperscript{18} is that tomosynthesis improves the results of digital mammograms only when two projections are taken or when combined with mammogram. Although it was early thought that tomosynthesis could reduce discomfort during breast examination using one single compression per breast, these results are indicative that a double compression will still be necessary. When it comes to combining tomosynthesis and mammograms the early machines obtained a mammogram and a tomosynthesis one after the other but the modern ones obtain mammograms (synthesis images) from the tomosynthesis thus reducing the dose of radiation. The FDA has approved a machine with these characteristics and a recent paper has confirmed that combining synthesis images and tomosynthesis improves the detection of breast cancer while reducing recalls yet as the authors indicate, clinical trials are necessary to assess the results of this combined technique during screening.\textsuperscript{27}

The good results from the negative odds ratio are promising for this test to be of diagnostic confirmation but they should be interpreted with caution because of the poor quality of the original studies whose level of evidence is not \(> II\) (cohort studies with methodological limitations); also they do not allow us to make valid recommendations. In general the studies had low statistic power which limits both the generalization of the results and the possibility of finding real differences between tomosynthesis and the digital mammogram. In addition the inadequate spectrum of patients and the double reference standard overestimated the validity of tomosynthesis while the significant loss of patients in several studies will mean bias whose significance could not be assessed.

External validity was limited by the broad heterogeneity of the studies in terms of population, machines and reading procedures. At least the maturity of the machines was not the only determinant of the heterogeneity of the results since it occurred both in studies conducted with commercial machines and with prototypes. We also believe that it is necessary to deepen the analysis of inter-observer variability since differences in the results based on the radiologist were seen yet the effect of his experience was different.\textsuperscript{15,21}

When it comes to the independence of the studies we must remember that most were supported by the industry. The formal aspects of informed consent were observed but we do not know if in the information from patients the specific recommendations on radiodiagnostics were taken into account.\textsuperscript{15,29}

Our review has the limitation of not having carried out a meta-analysis but it was not possible because of the studies high heterogeneity. On the other hand though we know that the probability of publishing works with positive results is greater we consider selection bias to be unlikely in our review since no limitations were established when it comes to the language of publication and the selection of the studies happened with predefined criteria by two independent researchers.

The clinical implication of these findings seems limited, because although its use as a diagnostic confirmation test could prevent the use of other tests before the biopsy and minimize the number of recalls on the operative plane it might well mean an increase in the time for examination and reading of the images. The impact on the screening process could be significant. The investment necessary to implement this technique would be high in each institution.
and be conditioned by the possibilities of evolution from the existing mammographs.

We do not have any studies on cost-effectiveness and the level of uncertainty is high, but an ECRI consensus report determined that its clinical impact can be very low, intermediate in its degree of use and high in both costs and processes. In sum the results suggest that tomosynthesis can be useful as a diagnostic confirmation test. With one projection it is not better than the digital mammogram with 2 yet results are way more promising with 2 projections and even more if combined with the mammogram. All of this taking into consideration that the level of evidence is still enough to be able to make any recommendations.

Ethical responsibilities

Protection of people and animals. Authors declare that for this investigation no experiments on human beings or animals were performed.

Data confidentiality. Authors declare that in this article there are no data from patients.

Right to privacy and informed consent. Authors declare that in this article there are no data from patients.

Funding

Collaboration agreement among the following institutions: Carlos III Institute of Health, Progress and Health Foundation of the Junta de Andalucía Office of Equality, Health, and Social Policies in the framework of developmental activities from the Spanish Network of Evaluation Agencies of Health Technologies and SNS Compensations funded by the Spanish Ministry of Health, Social Services and Equality.

Authors

1. Manager of the integrity of the study: FJGL.
2. Study Idea: ALLM.
3. Study Design: FJGL and ALLM.
4. Data Mining: FJGL and ALLM.
5. Data Analysis and Interpretation: FJGL and ALLM.
7. Reference Search: RIG.
8. Writing: FJGL and ALLM.
9. Critical review of the manuscript with intellectually relevant remarks: ALLM and RIG.
10. Approval of final version: FJGL, ALLM and RIG.

Conflict of interests

Authors declare no conflict of interests.

Acknowledgements

The authors wish to thank Dr. Marina Álvarez Benito, M.D. for the first draft of this study and the reviewers of this article.

References