Comprehensive analysis of clinical utility of three-dimensional ultrasound for benign and malignant breast masses

Jun Fu 1
Yanyan Li 2
Na Li 1
Zhanzhan Li 1

1Department of Oncology, 2Department of Nursing, Xiangya Hospital, Central South University, Changsha, Hunan Province 410008, China

Background: Three-dimensional (3-D) ultrasound is commonly used for screening breast cancer; however, the diagnostic accuracy of this method is unknown. Here, we performed a systematic search on the literature to assess the clinical utility of 3-D ultrasound in benign and malignant breast masses.

Materials and methods: We conducted searches in several online databases covering all publications prior to August 15, 2017. The bivariate random effects model was used to assess the overall sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and summary area under receiver operating curve (AUC) with their corresponding 95% CI.

Results: The overall sensitivity of 3-D ultrasound for diagnosing benign and malignant breast masses was 89% (95% CI, 83%–93%) and the specificity was 88% (95% CI, 83%–92%) with high heterogeneity (I² = 81.9; 95% CI, 74.4–89.3, P < 0.001). Other parameters used to assess efficacy included PLR (5.57; 95% CI, 3.73–8.31), NLR (0.18; 95% CI, 0.11–0.28), and DOR (31.33; 95% CI, 15.19–64.61). The use of a Fagan diagram with a pretest probability of 20% yields a post-test probability of 65% with a PLR of 7. True post-test probability was calculated at 3%, with an NLR of 0.13. The summary receiver operating characteristic curve was 0.94 (95% CI, 0.92–0.96), with no evidence of publication bias.

Conclusion: Three-dimensional ultrasound offers high sensitivity and specificity, with a high AUC, indicating a strong diagnostic value for detecting benign and malignant breast masses. Three-dimensional ultrasound may therefore represent an excellent option for secondary analysis of unclear breast lesions.

Keywords: breast cancer, diagnostic, meta-analysis, three-dimensional ultrasound

Introduction
Cancer has long been an important public health issue, resulting in significant physical and emotional damage to those affected. Since the early 1970s, breast cancer has been the most frequently diagnosed form of cancer in women, with incidence gradually increasing by 0.2%–0.8% each year, and the onset age tends to be more younger.1 Similar trends in breast cancer incidence are also seen in China, highlighting the importance of this issue in public health and the need for greater public attention.2

Preliminary screening for breast cancer typically involves imaging of susceptible breast tissue by methods such as computed tomography (CT), mammography X-ray examination, MRI, electrical capacitance tomography, and Doppler ultrasound with mammography X-ray examination. Despite the widespread use of Doppler ultrasound...
methods, in recent years, MRI, CT, and other imaging technologies have grown increasingly common; these methods offer significant advantages over traditional methods, although widespread adoption remains limited due to their higher cost.3

With the continued development and greater availability of color Doppler and contrast-enhanced ultrasound technologies, ultrasound has become the preferred method for breast tissue examination, suitable for differential diagnosis of physical properties of standard and cystic mass lesions. Three-dimensional (3-D) ultrasound works by reconstructing images from standard two-dimensional ultrasound. Besides providing an ultrasonic profile and cross-sectional information, the acquired coronal and stereoscopic images allow for detection and diagnosis of mammary lesions.3,5

The widespread adoption of these tools, combined with a greater understanding of enhancement patterns typical of breast lesions, has led to significant improvements in overall detection rates, including that of smaller breast lesions.6 A small number of studies have sought to assess the diagnostic abilities of 3-D ultrasound, as the overall sensitivity and specificity of this method remain unclear.7,8 Here, we conducted a comprehensive analysis of the public literature to evaluate the clinical utility of 3-D ultrasound for the diagnosis of benign and malignant breast masses among Chinese women.

Materials and methods
Search strategy
All relevant publications deposited in PubMed, Web of Science, Wanfang, and China National Knowledge Infrastructure prior to August 15, 2017, were included in this study. Relevant studies were identified using the following search terms: (breast cancer OR breast neoplasms OR breast carcinoma OR breast tumor OR breast) AND (ultrasound OR three-dimensional ultrasound OR three dimensions hypersound OR 3-D sonography) AND (sensitivity OR specificity OR diagnostic OR receiver operating curve OR ROC) in the title, abstract, or MeSH. Additional studies were identified by manually screening the reference lists of relevant reviews. Although we placed no restriction on language when performing the search, literature screening and full-text scanning were limited to English or Chinese. Two investigators independently performed the search; disagreements over inclusion were resolved by a third investigator.

Criteria for inclusion and exclusion
For a study to be included in this analysis, each of the following criteria had to be met: 1) the study evaluated the diagnostic accuracy of 3-D ultrasound for benign and malignant breast masses; 2) all patients were confirmed by a gold standard (pathological biopsy); 3) the study provided sufficient data for calculating the sensitivity and specificity including true positive (TP), false positive (FP), false negative (FN), and true negative (TN); and finally 4) the study was conducted on human subjects. Exclusion criteria included 1) experimentation studies, comments, reviews, letters, and conferences abstracts; and 2) studies with very small sample sizes (N<30). In cases of continuing or duplicate studies, only the most recent data were used.

Data extraction and quality assessment
Data extraction was performed by two independent investigators using a standard Excel sheet. For each study, the following information was extracted: the first author of the study, year of publication, type of examination equipment, frequency, number of lesions, mean age, sample size, four-folds (TP, FP, FN, TN), sensitivity, and specificity. A third investigator cross-referenced extraction information and revised the errors.

We used the quality assessment of diagnostic accuracy studies 2 (QUADAS-2) to assess the quality of included studies, based on four basic criteria: patient selection, index test, reference standard, and flow and timing.9 Each domain item is further broken down into several subitems designated as low risk, high risk, or unclear risk according to the study.

Statistical analyses
The threshold effect is one of the primary causes of heterogeneity within studies.10 For the current study, we evaluated the threshold effect using Spearman’s coefficient. In this study, the threshold effect does not exist (r=−0.358, P=0.059). Heterogeneity was assessed using chi-squared and I² statistical analyses. Outcomes significant at a level of P<0.05 or I²>50% were regarded as indicative of heterogeneity.11 The bivariate random effects model was used to assess the overall sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), diagnostic odds ratio (DOR), and summary area under receiver operating curve (AUC) with their respective 95% CI.12,13 Diagnostic ability was assessed via AUC. Here, an AUC=1 is indicative of a perfect distinguishing ability, while an AUC<0.5 signifies a poor diagnostic test. We used a Fagan diagram to evaluate pre- and post-test probabilities and Deek’s test for publication bias.14 Sensitivity analyses were also conducted using a standard leave-one-out method. All analyses were calculated using Stata 14.0 version.
(StataCorp LP, College Station, TX, USA), with \(P\)-values <0.05 considered statistically significant.

**Results**

**Study selection**

Our initial screening revealed 365 potential studies, from which 78 records were removed due to duplicated information. An additional 212 were excluded based on a manual screening of titles and abstracts, with most exclusions due to irrelevant topics and review articles. The remaining 75 articles were further screened through a reading of the full text. Twelve articles with unrelated diagnostic values, eight articles with insufficient data, eight studies with duplicates data, 17 case reports, and 11 reviews, comments, and letters were excluded. After all exclusions had been made, we were left with 19 studies that met all of our quantitative and qualitative inclusion criteria. A flowchart outlining this screening process is shown in Figure 1.

**General characteristics of studies and quality assessment**

All included studies were published between 2007 and 2014. The mean age ranged from 30 to 56 years. There was a total of 2,165 patients; sample sizes ranged from 51 to 242 patients. Most studies used a GE Voluson ultrasound for examination, run at 5–12 MHz. The total number of lesions detected across all studies was 2,214. Sensitivity scores ranged from 62% to 99%, with a specificity of 71% to 99%. A summary of the patient cohorts included in this study is shown in Table 1.

**Pooled results**

Nineteen studies reported sensitivity and specificity data. The heterogeneity within studies was high (\(I^2=85.0; 95\%\) CI, 79.2–90.8, \(P<0.001\)). The overall sensitivity and diagnostic accuracy of 3-D ultrasound for breast cancer was 89% (95\% CI, 83%–93%; Figure 2). Overall specificity was 88% (95\% CI, 83%–92%; Figure 3) with high heterogeneity (\(I^2=81.9\); 95\% CI, 72.3–90.8, \(P<0.001\)).
95% CI, 74.4–89.3, P < 0.001). Other parameters used to assess efficacy included PLR (5.57; 95% CI, 3.73–8.31), NLR (0.18; 95% CI, 0.11–0.28), and DOR (31.33; 95% CI, 15.19–64.61).

The use of a Fagan diagram with a pretest probability of 20% yields a post-test probability of 65% with a PLR of 7 (Figure 5). True post-test probability was calculated at 3%, with an NLR of 0.13. The summary receiver operating characteristic curve was 0.94 (95% CI, 0.92–0.96; Figure 4), indicative of the high diagnostic accuracy of 3-D ultrasound for breast cancer.

### Sensitivity analyses and publication bias

To examine the stability of our pooled results, we also conducted sensitivity analyses using a standard leave-one-out method. As shown in Figure 6, the sensitivity ranged from

---

**Table 1** General characteristics of studies included in the meta-analysis

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Machine</th>
<th>MHz</th>
<th>Lesions (n)</th>
<th>Mean age</th>
<th>Sample size (case/control)</th>
<th>TP</th>
<th>FP</th>
<th>FN</th>
<th>TN</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang and Hu15</td>
<td>2007</td>
<td>GE Voluson E8</td>
<td>6–12</td>
<td>130</td>
<td>40</td>
<td>107</td>
<td>61</td>
<td>1</td>
<td>2</td>
<td>66</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>Cheng et al16</td>
<td>2008</td>
<td>Accuvix XQ</td>
<td>10</td>
<td>51</td>
<td>46</td>
<td>51</td>
<td>25</td>
<td>1</td>
<td>1</td>
<td>24</td>
<td>96</td>
<td>96</td>
</tr>
<tr>
<td>Hong et al17</td>
<td>2009</td>
<td>GE Voluson</td>
<td>8–12</td>
<td>129</td>
<td>49</td>
<td>129</td>
<td>47</td>
<td>6</td>
<td>12</td>
<td>64</td>
<td>80</td>
<td>91</td>
</tr>
<tr>
<td>An et al18</td>
<td>2010</td>
<td>Accuvix Q8</td>
<td>8–15</td>
<td>50</td>
<td>45</td>
<td>83</td>
<td>27</td>
<td>12</td>
<td>6</td>
<td>38</td>
<td>82</td>
<td>76</td>
</tr>
<tr>
<td>Xu and Cao19</td>
<td>2010</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>160</td>
<td>48</td>
<td>160</td>
<td>65</td>
<td>15</td>
<td>13</td>
<td>78</td>
<td>83</td>
<td>84</td>
</tr>
<tr>
<td>Zhou et al20</td>
<td>2011</td>
<td>Philips</td>
<td>5–13</td>
<td>142</td>
<td>46</td>
<td>142</td>
<td>56</td>
<td>10</td>
<td>5</td>
<td>71</td>
<td>92</td>
<td>88</td>
</tr>
<tr>
<td>Li et al21</td>
<td>2012</td>
<td>LOGIQ9</td>
<td>–</td>
<td>152</td>
<td>49</td>
<td>126</td>
<td>56</td>
<td>17</td>
<td>6</td>
<td>73</td>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>Huang and Xu22</td>
<td>2013</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>60</td>
<td>44</td>
<td>60</td>
<td>40</td>
<td>2</td>
<td>1</td>
<td>28</td>
<td>98</td>
<td>93</td>
</tr>
<tr>
<td>Ma et al23</td>
<td>2013</td>
<td>LOGIQ9</td>
<td>5–12</td>
<td>85</td>
<td>–</td>
<td>85</td>
<td>41</td>
<td>11</td>
<td>12</td>
<td>36</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>Zhang et al24</td>
<td>2013</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>79</td>
<td>55</td>
<td>79</td>
<td>39</td>
<td>6</td>
<td>2</td>
<td>34</td>
<td>95</td>
<td>85</td>
</tr>
<tr>
<td>Yang et al25</td>
<td>2013</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>112</td>
<td>56</td>
<td>112</td>
<td>62</td>
<td>16</td>
<td>7</td>
<td>40</td>
<td>90</td>
<td>71</td>
</tr>
<tr>
<td>Xie et al26</td>
<td>2014</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>201</td>
<td>54</td>
<td>201</td>
<td>83</td>
<td>13</td>
<td>15</td>
<td>137</td>
<td>85</td>
<td>91</td>
</tr>
<tr>
<td>Hu et al27</td>
<td>2014</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>107</td>
<td>40</td>
<td>107</td>
<td>29</td>
<td>18</td>
<td>7</td>
<td>73</td>
<td>81</td>
<td>80</td>
</tr>
<tr>
<td>Dou28</td>
<td>2014</td>
<td>Accuvix XQ</td>
<td>10</td>
<td>150</td>
<td>45</td>
<td>150</td>
<td>12</td>
<td>6</td>
<td>14</td>
<td>118</td>
<td>46</td>
<td>95</td>
</tr>
<tr>
<td>Sun et al29</td>
<td>2014</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>115</td>
<td>39/46</td>
<td>115</td>
<td>60</td>
<td>5</td>
<td>4</td>
<td>46</td>
<td>94</td>
<td>90</td>
</tr>
<tr>
<td>Xu and Xue30</td>
<td>2014</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>242</td>
<td>44</td>
<td>242</td>
<td>119</td>
<td>5</td>
<td>2</td>
<td>174</td>
<td>98</td>
<td>97</td>
</tr>
<tr>
<td>Zhang et al31</td>
<td>2014</td>
<td>GE Voluson</td>
<td>5–12</td>
<td>90</td>
<td>56</td>
<td>90</td>
<td>44</td>
<td>14</td>
<td>11</td>
<td>21</td>
<td>80</td>
<td>60</td>
</tr>
</tbody>
</table>

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive; –, not applicable.
Three-dimensional ultrasound and breast cancer

~88% to 96%, with a specificity of 86% to 98%, indicative of high stability. Assessment of publication bias was performed using Deek’s test, with no evidence of bias between studies ($t = 0.290; P = 0.781$; Figure 7).

**Discussion**

Our analyses revealed an overall sensitivity of 3-D ultrasound of 89%, specificity of 88%, and an AUC of 0.94. These observations are indicative of the high diagnostic ability of 3-D ultrasound for both benign and malignant breast masses. Our pooled PLR was 31.33; thus, patients with malignant breast masses have a nearly 4.5-fold higher chance of a positive 3-D ultrasound compared with those without a benign breast mass. The NLR was 0.13, meaning the probability of a missed diagnosis was 13% if the 3-D ultrasound was negative. Current guidelines suggest a PLR $>10$ and NLR $<0.1$ as a baseline for high diagnostic ability.

Two-dimensional ultrasound remains the most common method for routine breast examination. Unlike 3-D approaches, two-dimensional ultrasound only produces slice images, meaning clinicians must on their own construct a series of 3-D images to fully visualize the area. Such a limitation greatly increases the likelihood that some lesions will be missed, particularly in cases where highly skilled technicians are not available. Moreover, it is extremely difficult for clinicians to locate lesions observed in a previous examination, which serves as a significant disadvantage for long-term follow-up. In moving to 3-D ultrasound, doctors can pinpoint the exact location, size, and boundaries of a mass. In addition, other images of the interface, such as breast coronal images, can be obtained by 3-D ultrasound, thereby improving the accuracy of the diagnosis.
The ability to evaluate the size, range, form, and malignancy of a breast lump is an important application of ultrasound. Zhang et al found the sensitivity of 3-D ultrasound to be 95.8%, a significant improvement over traditional two-dimensional scan (89.6%). These 3-D scans also offer significantly better accuracy in terms of lesion locations, especially for the lesions found in the cavity of breast ducts. The molybdenum target remains the “gold standard” for lesional imaging, although the accuracy of this method is limited in dense breasts, giving ultrasound a clear advantage in these cases. The molybdenum target remains the “gold standard” for lesional imaging, although the accuracy of this method is limited in dense breasts, giving ultrasound a clear advantage in these cases.

Such a consideration may be of particular importance in China, as the breasts of Chinese women are typically smaller and denser than those of American and European women. Used together, the combination of 3-D ultrasound and a molybdenum target can be used to significantly improve patient diagnosis in the clinical setting. Previous studies reported that the 3-D images of breast cancer are characterized by unclear boundary, convergent feature, and abundant blood flow. The benign lesions in three-dimensional imaging were presented in the form of clear boundaries, regular morphology and the echogenicity was homogeneous and blood flow signal was not rich, and there was no clear boundary between lesions and adjacent tissues. No mass was found for this kind of lesions under the radiographs with

Figure 5 Fagan diagram evaluating the overall diagnostic value of three-dimensional ultrasound for breast cancer.

Abbreviations: pos, positive; neg, negative; prob, probability; LR, likelihood ratio.
Figure 6 Sensitivity and pooled sensitivity and specificity.

Figure 7 Deek’s plot of publication bias.

Abbreviations: sqrt, square root; ESS, effective sample sizes.
molybdenum target. These characteristics can be helpful in differentiating malignant and benign lesions.

Our study had several limitations. First, all of the included studies consisted of patients with benign and malignant tumors, with some studies not providing clear descriptions regarding blinding methods, which may lower the study quality. Second, the diagnostic criteria for benign or malignant tumors differed between studies, with some studies treating specific signs as a benign diagnosis, thereby lowering the reported diagnostic efficiency of 3-D ultrasound. Third, studies were conducted using different examination machines with different frequencies, which may underestimate or overestimate the diagnostic accuracy for breast cancer. Finally, patient cohorts differed across studies, which may confound study outcome. More studies will be necessary to fully evaluate the diagnostic value of 3-D ultrasound for benign and malignant breast masses.

Conclusion
Taken together, our results indicated that 3-D ultrasound has a high clinical value for diagnosing benign and malignant breast masses. The use of 3-D ultrasound therefore offers a simple and inexpensive method for diagnosing uncertain breast lesions. Early diagnosis and early treatment of secondary prevention of tumor are the key to improve the prognosis of breast cancer patients. The application value of 3-D ultrasound examination in mammary gland disease is certain, has simple, noninvasive, good repeatability, and is not affected by the density degree of mammary gland and examination position. The confirmation of benign or malignant breast masses assists the clinical decision.

Disclosure
The authors report no conflicts of interest in this work.

References