Contrast-enhanced ultrasound for differential diagnosis of malignant and benign ovarian tumors: systematic review and meta-analysis

X. MA*,†#, Y. ZHAO*,#, W. LING‡, H. ZHUO†, H. JIA* and P. LI†

*Cancer Center, State Key Laboratory of Biotherapy, West China Hospital of Sichuan University, Chengdu, Sichuan, PR China; †Department of Oncology, West China Hospital of Sichuan University, Chengdu, Sichuan, PR China; ‡Department of Ultrasound, West China Hospital of Sichuan University, Chengdu, Sichuan, PR China

KEYWORDS: adnexal nodules; contrast-enhanced ultrasound; diagnosis; ovarian cancer

ABSTRACT

Objective To assess the performance of contrast-enhanced ultrasound (CE-US) in the differential diagnosis of malignant and benign ovarian tumors.

Methods We conducted a comprehensive literature search of PubMed and EMBASE to identify published articles evaluating the diagnostic potential of CE-US for the differentiation of benign and malignant ovarian tumors. Inclusion criteria were: (1) the study assessed the accuracy (or sensitivity and specificity) of CE-US for diagnosis of benign and malignant ovarian tumors; (2) it used surgery and histopathology as the reference standard for distinguishing between benign and malignant tumors; (3) it included data allowing construction of a 2×2 contingency table for true- and false-positives and negatives. We present summary sensitivity, specificity, diagnostic odds ratio (OR) and areas under the summary receiver–operating characteristics curves (AUCs).

Results Preliminary screening identified 103 papers, of which 11 fulfilled our predefined inclusion criteria and underwent final analysis. The pooled sensitivity and specificity of CE-US for diagnosis of benign and malignant ovarian tumors were 93% (95% CI, 89–96%) and 95% (95% CI, 92–96%), respectively. The pooled diagnostic OR was 171.2 (95% CI, 65.9–444.6) and the AUC was 0.98. I² values of sensitivity, specificity and diagnostic OR were 38.3%, 31.7% and 48.4%, respectively, all indicating moderate heterogeneity.

Conclusions The evidence from available studies suggests CE-US is useful for discriminating between benign and malignant ovarian tumors; however, further studies are needed to examine whether CE-US has improved diagnostic test accuracy compared with that of standard two-dimensional Doppler sonography. Copyright © 2015 ISUOG. Published by John Wiley & Sons Ltd.

INTRODUCTION

According to global cancer statistics, in 2008 about 225,500 people had ovarian cancer and 140,200 patients died from the disease. Globally, ovarian cancer is the sixth most common cancer among women, accounting for 4% of all female cancers worldwide. Despite advances in surgery, chemotherapy and radiotherapy, survival rates have not increased significantly since then, the 5-year survival rate being around 40%. Moreover, about two-thirds of patients are diagnosed with advanced disease, which is why ovarian cancer remains the leading cause of death from gynecological malignancy in the USA; the earlier the diagnosis, the better the therapeutic strategies and the higher the survival rates. Thus, new technologies for detection of early-stage ovarian cancer are needed urgently.

Contrast-enhanced ultrasound (CE-US), an ultrasound imaging technique which uses a microbubble contrast agent, such as SonoVue or Levovist, has been proposed to improve the diagnostic test accuracy of ultrasonography in examining ovarian tumors. The contrast medium is injected intravenously into the circulation, where the microbubbles provide a harmonic non-linear signal when insonated, whereas the surrounding tissues provide a weak signal, thereby increasing the contrast and improving the imaging quality. Previous evidence has shown that CE-US signaling patterns differ between benign and malignant lesions.
The objective of this systematic review and meta-analysis was to identify, appraise and summarize the available evidence on the diagnostic test accuracy of CE-US in distinguishing between benign and malignant ovarian tumors.

METHODS

Protocol and registration

Our meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations5. It was not, however, registered.

Eligibility criteria

The inclusion criteria for articles assessing the performance of CE-US in the diagnosis of benign and malignant ovarian tumors were set before the literature search was performed. Studies were selected for inclusion in the meta-analysis if they met all of the following criteria: (1) they assessed the accuracy of CE-US in the diagnosis of benign and malignant ovarian tumors; (2) they used surgery and histopathology as the reference standard for distinguishing between benign and malignant tumors; (3) they included data allowing construction of a 2×2 contingency table for true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN).

Literature search

We searched PubMed and EMBASE for studies published up to 8 December 2014. The search terms used were combinations of the relevant medical subject heading (MeSH) terms, keywords and word variants of ‘ovarian’, ‘neoplasm’, ‘contrast-enhanced ultrasound’, ‘sonovue’ and ‘levovist’. The full search strategy is available in Appendix S1. We also carried out a manual search by searching the reference lists of relevant articles to expand our included studies.

Study selection

Abstracts were reviewed for eligibility according to the predefined criteria by two independent observers (Y.Z. and X.M.). Disagreements were resolved by consulting a third observer (B.Z.). Full-text papers considered potentially eligible for inclusion were then retrieved and the two observers (Y.Z. and X.M.) evaluated their eligibility independently. Disagreements were resolved by consulting the third observer (B.Z.).

Data collection process and data items

Relevant data about the characteristics of the study were extracted by two observers independently (Y.Z. and X.M.); any cases of disagreement were checked by another observer (W.L.) independently and resolved by consensus. The following items were extracted according to a fixed protocol: author, year of publication, country, description of original study population (number of patients, mean age), number of ovarian lesions, ultrasound modality, contrast agents and ultrasound equipment.

Diagnostic 2×2 contingency tables for TP, FP, TN and FN could be extracted or derived from reported sensitivity, specificity and positive and negative predictive values.

Statistical methods

A random-effects model was used in this meta-analysis. Overall pooled sensitivities, specificities and positive
Contrast-enhanced sonography of ovarian tumors

Table 1 Summary table of the meta-analysis

<table>
<thead>
<tr>
<th>Reference</th>
<th>Country</th>
<th>Study type</th>
<th>Mean age (years)</th>
<th>Patients</th>
<th>Masses</th>
<th>Sens. (%)</th>
<th>Spec. (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>TP (n)</th>
<th>FP (n)</th>
<th>FN (n)</th>
<th>TN (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang18 (2014)</td>
<td>China</td>
<td>Retro</td>
<td>52</td>
<td>120</td>
<td>120</td>
<td>CE-US</td>
<td>Sens.: 93.5</td>
<td>Spec.: 97.2</td>
<td>PPV: 95.6</td>
<td>NPV: 95.9</td>
<td>43</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Hu17 (2014)</td>
<td>China</td>
<td>Retro</td>
<td>40.6</td>
<td>57</td>
<td>57</td>
<td>2D- and 3D-CE-US</td>
<td>Sens.: 60</td>
<td>Spec.: 54.6</td>
<td>PPV: 91.3</td>
<td>NPV: 84.8</td>
<td>50</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>Yang19 (2013)</td>
<td>China</td>
<td>Retro</td>
<td>44.4</td>
<td>106</td>
<td>106</td>
<td>CE-US</td>
<td>Sens.: 93.3</td>
<td>Spec.: 90.3</td>
<td>PPV: 97.9</td>
<td>NPV: 100</td>
<td>70</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Xiang15 (2013)</td>
<td>China</td>
<td>Prosp</td>
<td>43.2</td>
<td>47</td>
<td>51</td>
<td>3D-CE-US</td>
<td>Sens.: 100</td>
<td>Spec.: 98</td>
<td>PPV: 97.1</td>
<td>NPV: 100</td>
<td>94</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>Huchon14 (2012)</td>
<td>France</td>
<td>Prosp</td>
<td>45.8</td>
<td>99</td>
<td>99</td>
<td>CE-US</td>
<td>Sens.: 95.7</td>
<td>Spec.: 94.7</td>
<td>PPV: 97.1</td>
<td>NPV: 98.8</td>
<td>79</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Zhou13 (2009)</td>
<td>China</td>
<td>Retro</td>
<td>42</td>
<td>65</td>
<td>65</td>
<td>Color Doppler CE-US</td>
<td>Sens.: 96.7</td>
<td>Spec.: 97</td>
<td>PPV: 96.7</td>
<td>NPV: 97.1</td>
<td>29</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Ordén16 (2003)</td>
<td>Finland</td>
<td>Retro</td>
<td>49.4</td>
<td>70</td>
<td>70</td>
<td>Color Doppler CE-US</td>
<td>Sens.: 93</td>
<td>Spec.: 93</td>
<td>PPV: 93</td>
<td>NPV: 93</td>
<td>13</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Kurjak10 (2001)</td>
<td>Croatia</td>
<td>Retro</td>
<td>54</td>
<td>89</td>
<td>NR</td>
<td>3D power Doppler CE-US</td>
<td>Sens.: 100</td>
<td>Spec.: 99.9</td>
<td>PPV: 99.9</td>
<td>NPV: 100</td>
<td>12</td>
<td>12</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 1 Only the first author of each study is given. Four studies compared CE-US with conventional US: FN, false negative; FP, false positive; PPV, positive predictive value; NPV, negative predictive value; NR, not reported; PPV, positive predictive value; Retrospective: Sens., sensitivity; Spec., specificity; TN, true negative; TP, true positive; US, ultrasound.

Risk of bias in individual studies

To assess the methodological quality and applicability of the studies included in our meta-analysis, we used the revised tool for the quality assessment of diagnostic accuracy studies (QUADAS-2). Two independent observers (Y.Z. and X.M.) performed the quality assessment, rating the items of the four key domains (patient selection, index test, reference standard, and flow and timing) as: 'yes (high quality)' if reported; 'no (low quality)' if not reported; or 'unclear' if no adequate information was provided from which to make an accurate judgment.

RESULTS

Study selection

The screening and inclusion process for the meta-analysis is presented in Figure 1. The initial search included a total of 110 articles, of which seven were duplicates and therefore removed. After screening titles and abstracts, a further 55 were excluded. After examination of the full text of the remaining 48 candidate studies, a further 34 were excluded, leaving 14 articles considered as being eligible for the systematic review. Three of these were not eligible for meta-analysis due to missing sensitivity, specificity, accuracy or correlation values; thus, 11 studies were eligible for the meta-analysis according to our inclusion criteria.

Study characteristics

The characteristics of the 11 included studies are summarized in Table 1. All were published between September 2000 and April 2014. Three12,14,15 were prospective studies and eight9−11,13,16−19 were retrospective. The number of patients in each study varied from 20 to 120 (mean, 72.5). All studies used some form of CE-US, the modality being generally color Doppler CE-US, power (LR+) and negative (LR−) likelihood ratios (with corresponding 95% CIs) were calculated from the TP, FP, FN and TN from each study, to indicate the accuracy of CE-US in the diagnosis of benign and malignant tumors. In addition, the summary receiver–operating characteristics (sROC) curve was constructed, as described by Moses et al.6. The Youden’s index was calculated as sensitivity + specificity – 1. The heterogeneity of the pooled studies was assessed by the inconsistency index (I²). P > 50% indicates significant heterogeneity. Publication bias was assessed by funnel plot asymmetry test and quantitatively by Deeks’ test, P > 0.05 suggesting no significant publication bias. All statistical analyses were performed using MetaDiSc statistical software version 1.417 (Hospital Universitario, Madrid, Spain) and STATA 12.0 (STATA, College Station, TX, USA).
Doppler CE-US or three-dimensional (3D) CE-US, in addition to conventional two-dimensional (2D) color and power Doppler ultrasonography. The microbubble contrast agent used was SonoVue or Levovist. The diagnostic accuracy of CE-US was assessed with respect to pathology in all studies.

Accuracy of CE-US for differential diagnosis between malignant and benign ovarian tumors

Pooled estimates of CE-US sensitivity and specificity for the diagnosis of benign and malignant ovarian tumors were 93% (95% CI, 89–96%) and 95% (95% CI, 92–96%), respectively (Figure 2). LR+ was 14.37 (95% CI, 9.07–22.78) and LR– was 0.10 (95% CI, 0.05–0.19). The diagnostic odds ratio (DOR) was 171.2 (95% CI, 65.9–444.6). The symmetrical sROC curve for the diagnostic value of CE-US is given in Figure 3; the area under the curve (AUC) was 0.98.

The $I^2$ values for sensitivity, specificity and DOR were 38.3%, 31.7% and 48.4%, respectively, indicating moderate heterogeneity for all three. The conclusions from three studies\textsuperscript{20–22} that examined CE-US but were not included in the quantitative analysis were also assessed (data not shown); all three suggested that CE-US is useful for discriminating benign from malignant ovarian tumors.

Comparison of CE-US and conventional US for differential diagnosis between malignant and benign ovarian tumors

Four studies\textsuperscript{10,13,14,19} included sensitivity, specificity, accuracy or correlation values of conventional US and thus we were able to undertake a comparison of CE-US and conventional US for the differential diagnosis of malignant and benign ovarian tumors. Figure 4 presents pooled estimates of sensitivity and specificity, and Figure 5
Contrast-enhanced sonography of ovarian tumors

Figure 3 Summary receiver–operating characteristics curve for contrast-enhanced ultrasound in the differential diagnosis of benign and malignant ovarian tumors. Individual studies are plotted, with sizes of circles reflecting relative sizes of study cohorts. Area under the curve (AUC) = 0.9814; SE (AUC) = 0.008; Q* = 0.9398; SE (Q*) = 0.0157.

presents the symmetrical sROC curve for conventional US. The sensitivity, specificity, LR+, LR−, accuracy of CE-US and Youden index in the diagnosis of benign and malignant ovarian tumors were 94%, 94%, 15.58, 0.07, 94% and 0.88, while those of conventional US were 83%, 92%, 8.09, 0.19, 87% and 0.75.

Figure 4 Forest plots of study-specific estimates of sensitivity and specificity of conventional ultrasound in differentiation of benign from malignant ovarian tumors.

Figure 5 Symmetrical summary receiver–operating characteristics curves for conventional ultrasound in the differential diagnosis of benign and malignant ovarian tumors. Individual studies are plotted, with sizes of circles reflecting relative sizes of study cohorts. Area under the curve (AUC) = 0.9084; SE (AUC) = 0.029; Q* = 0.8404; SE (Q*) = 0.0319.

Assessment of bias

Deeks’ funnel plot asymmetry test was employed to examine the publication bias (Figure 6). There was no significant publication bias for the DOR of CE-US in the differential diagnosis of malignant and benign ovarian tumors.

<table>
<thead>
<tr>
<th>Study</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yang19</td>
<td>0.92 (0.83–0.97)</td>
<td>0.68 (0.49–0.83)</td>
</tr>
<tr>
<td>Huchon14</td>
<td>0.55 (0.23–0.83)</td>
<td>0.94 (0.87–0.98)</td>
</tr>
<tr>
<td>Zhou13</td>
<td>0.73 (0.54–0.88)</td>
<td>0.86 (0.70–0.95)</td>
</tr>
<tr>
<td>Kurjak10</td>
<td>0.80 (0.61–0.92)</td>
<td>0.95 (0.91–0.97)</td>
</tr>
</tbody>
</table>

Pooled sensitivity = 0.83 (0.76 to 0.89)  
Pooled specificity = 0.92 (0.88 to 0.94)  
Chi-square = 11.90; df = 3 (P = 0.0078)  
Inconsistency (I²) = 74.8%  
Chi-square = 20.39; df = 3 (P = 0.0001)  
Inconsistency (I²) = 85.3%
tumors \( (P = 0.27) \). The risk of bias and applicability concerns of included studies based on QUADAS-2 are summarized in Figure 7.

DISCUSSION

This meta-analysis included 11 studies which evaluated the diagnostic potential of CE-US for the differentiation between benign and malignant ovarian tumors and which met the inclusion criteria and provided sufficient data for this meta-analysis. The sensitivity and specificity were 93% and 95%, respectively, the DOR was 171.2 and the AUC was 0.98. There was moderate heterogeneity among the studies. This suggests that CE-US has high discriminatory accuracy in the diagnosis of benign and malignant ovarian tumors.

Our comparison between CE-US and conventional US for the differential diagnosis of malignant and benign ovarian tumors suggests that CE-US has higher accuracy (sensitivity, 94% vs 83%; specificity, 94% vs 92%; LR+, 15.58 vs 8.09; LR−, 0.07 vs 0.19; accuracy, 94% vs 87%; Youden index, 0.88 vs 0.75). However, with respect to clinical application, more multicenter studies are needed to affirm the superiority of CE-US over standard 2D Doppler sonography in the differentiation between these tumors.

Scoring systems have also been applied in the differential diagnosis of malignant and benign ovarian tumors. Xiang et al.\textsuperscript{15} generated a scoring system to differentiate benign small adnexal masses from malignant ones, based on the characteristics of these lesions on 3D-CE-US imaging; they evaluated surface, wall thickness, inner wall structures, septa and contrast enhancement of the masses, relationship with surrounding tissues and ascites. Based on the ROC curves generated from their 3D-CE-US scoring system, a cut-off of 8 was defined, with scores \( > 8 \) suggesting malignancy. Use of scoring systems in the diagnosis of benign and malignant ovarian tumors could reduce the influence of subjective judgment by sonographers.

CE-US is generally performed with 3D sonography, color Doppler or power Doppler. Among these three modalities, 3D Doppler sonography has been reported to reduce the FP rate in cystic-solid and solid ovarian masses\textsuperscript{23}. In addition, 3D-CE-US has great potential in the early detection of ovarian cancer; according to Xiang et al.\textsuperscript{15}, 3D-CE-US might show not only the large feeding blood vessels and their distribution in the lesion but also the septa and papillary structures on the inner walls of the lesion that are generally detected with difficulty. However, whether 3D Doppler sonography can improve the diagnostic performance of CE-US needs to be further researched.

There were several limitations in this study. First, we did not evaluate intraobserver or interobserver variability; it is important to do so because ultrasonography is an operator-dependent procedure. Second, the limited number of cases in combination with the wide variety of pathological types of ovarian tumor (such as ovarian theca cell tumor, yolk sac tumor and ovarian fibroma) may have caused bias due to the heterogeneity between studies.

In conclusion, the available evidence suggests that CE-US is useful for discriminating between benign and

![Figure 6](image-url)  
Figure 6 Deek's funnel plot asymmetry test for publication bias for the 11 studies included in the meta-analysis, with regression line. Test for asymmetry, \( P = 0.27 \). ESS, effective sample size.

![Figure 7](image-url)  
Figure 7 Summary of methodological quality and applicability of studies included in the meta-analysis, according to quality assessment of diagnostic accuracy studies (QUADAS-2)\textsuperscript{8} tool. Proportions of studies with low [■], high [■] or unclear [□] risk of bias (a) or concerns regarding applicability (b) are shown.
malignant ovarian tumors and could potentially be applied in clinical practice. More multicenter studies are needed to confirm our results and to assess further whether CE-US can improve the diagnostic test accuracy over conventional 2D color and power Doppler sonography.

REFERENCES

SUPPORTING INFORMATION ON THE INTERNET
The following supporting information may be found in the online version of this article:

Appendix S1 Search strategy for systematic review