Meta-analysis of contrast-enhanced ultrasonography for the detection of gallbladder carcinoma.

Wei Wang1*, Yang Fei2*, Feng Wang1

1Department of General Surgery, Bayi hospital affiliated Nanjing University of Chinese medicine/ the 81st Hospital of PLA, 2Nanjing University School of Medicine, Nanjing, China
*the authors shared the first authorship

Abstract

Aim: To systematically evaluate the accuracy of contrast-enhanced ultrasonography (CEUS) for the detection of gallbladder carcinoma. Material and method: Relevant studies were identified searching PubMed, Embase, Elsevier, the CNKI (China National Knowledge Infrastructure) database and the Cochrane Trial Register searches until August 2015. Patient clinical characteristics, sensitivity and specificity were extracted. The summary receiver operating characteristic (ROC) curve was used to examine the accuracy of CEUS. A meta-analysis was performed to evaluate the clinical utility in the diagnosis of gallbladder carcinoma. Sensitivity analysis was performed after omitting outliers identified in a bivariate boxplot and publication bias was assessed with Egger testing. Results: From 89 citations, 16 were included in the meta-analysis with a total of 1673 lesions. We detected heterogeneity between studies and evidence of publication bias. The methodological quality was moderate. The pooled weighted sensitivity with a corresponding 95% confidence interval (CI) was 0.92 (95%CI: 0.90, 0.94), the specificity was 0.91 (95%CI: 0.89, 0.93), the positive likelihood ratio was 10.01 (95%CI: 7.02, 14.29), the negative likelihood ratio was 0.10 (95%CI: 0.07, 0.14), and the diagnostic odds ratio was 123.02 (95%CI: 78.40, 193.03). The area under the ROC curve was 0.9689 (95%CI: 0.9376, 0.9879).

Conclusions: CEUS is a reliable, non-invasive, and no-radiation-exposure imaging modality with a high sensitivity and specificity for detection of gallbladder carcinoma. Nonetheless, it should be applied cautiously, and large scale, well-designed trials are necessary to assess its clinical value.

Keywords: contrast-enhanced ultrasonography, gallbladder carcinoma, meta-analysis

Introduction

Gallbladder carcinoma is the most common malignancy of the biliary system [1,2]. Early diagnosis of gallbladder carcinoma remains difficult because of the non-specificity of its clinical manifestations. When a diagnosis is made, radical surgical resection often cannot be performed because of frequent invasions into important circumferential structures such as the hepatic artery or the portal vein [1-3]. So, accurate preoperative evaluation of gallbladder cancer is essential for curative resection.

Ultrasonography, computer tomography (CT), and magnetic resonance imaging (MRI) are considered to be the best modalities for the diagnosis of gallbladder disease [4-6]. But the preoperative diagnosis of gallbladder cancer can be often difficult.

Contrast-enhanced ultrasonography (CEUS) is a new technique that has been reported to be useful for gallbladder cancer diagnosis [7]. Nowadays, more and more studies are focusing on the use of CEUS in the diagnosis of gallbladder carcinoma and our meta-analysis provides summaries of the results of relevant studies, estimates of the average diagnostic accuracy of CEUS, the uncertainty of this average, and the variability of the study findings around the estimates.
Material and method

**Literature Search**

PubMed, Embase, Elsevier, CNKI (China National Knowledge Infrastructure) database and Cochrane Trial Register searches were performed to identify all the eligible papers. The search terms were used as the following: contrast-enhanced ultrasonography or ultrasound, gallbladder, cancer carcinoma or tumor. The publication languages were restricted to English and Chinese. Moreover, potentially relevant studies were evaluated by reviewing the titles and abstracts, and studies matching the criteria were carefully retrieved. If more than one study was published using the same data, only the study with a larger population was included. The literature search was updated on August, 2015. This systematic review was planned, conducted, and reported in adherence to the standards of quality for reporting meta-analysis.

**Inclusion and exclusion criteria**

Studies were included in the meta-analysis if they met all of the following criteria: 1) full text original article; 2) human study; 3) evaluation of CEUS for the gallbladder carcinoma; 4) inclusion of at least 20 patients; 5) published in English or Chinese; 6) included an accepted reference method, using specimens obtained from surgery.

Exclusion criteria were 1) no evaluation of the value of CEUS for the diagnosis of gallbladder carcinoma; 2) samples less than 20 patients; 3) review article (including meta-analyses), corresponding letter or editorial not reporting original data; 4) published in abstract form only; 5) published more than once; 6) sufficient data including gender, age, lesion size and imaging modality for gallbladder carcinoma were not provided.

**Quality assessment of studies**

The methodological quality of the included studies was assessed independently by two observers using the revised tool for Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) instrument, a quality assessment tool specifically developed for systematic reviews of diagnostic accuracy studies [8]. The full QUADAS-2 tool consists of four domains: patient selection, index test, reference standard, and flow and timing. Each domain was assessed in terms of the risk of bias according to the signaling questions, and the first three domains were judged in terms of concerns regarding applicability. Each question was scored “yes” if reported, “no” if not reported, or “unclear” if there was inadequate information in the article to make a judgment. To resolve disagreement between the two reviewers, a third reviewer assessed all of the involved items. The majority opinion was used for the analysis.

**Data extraction**

Three researchers extracted data from each study by using a structured sheet and entered the data into a database. The following data were extracted from each study: the first author’s last name, publication year, source journal, research design, study duration, retention and dropout rate, surgical methods, patient’s sex and age, sample size. For each study, values for true-positive (TP), false-positive, true-negative (TN), false-negative, sensitivity (Sen), specificity (Spe), positive likelihood ratio (PLR) and negative likelihood ratio (NLR) results for the detection of patients were extracted too, and 2×2 contingency tables were constructed.

**Risk of bias assessment**

Assessment of methodological quality was performed using the risk of bias assessment tool by the Cochrane Collaboration indicating the following bias domains: selection bias (random sequence generation, allocation concealment), performance/detection bias (blinding of participants and personnel/blinding of outcome assessment), attrition bias (incomplete data outcome), reporting bias (selective reporting), and other bias.

In addition, a bivariate box plot with Egger testing was used to assess the distributional properties of sensitivity versus specificity and for identifying possible outliers. After omitting these outliers and according to the results of the subgroups analysis, sensitivity analysis was performed and the change in heterogeneity was observed.

**Statistical Analysis**

The Q statistic of the Chi-square value test and the inconsistency index (I-squared, I²) were used to estimate the heterogeneity of the individual studies using the STATA software 11.0 (Stata Corporation, Texas, USA). I² values were interpreted according to the proposal of Higgins et al [9], with heterogeneity determined as either low (I²≤25%), medium (25% <I²≤50%) or high (50% <I²≤75%). In this study, meta-regression was used to explore such heterogeneity by relating the accuracy measurement to study level covariates.

If notable heterogeneities were detected, the test performance was summarized by using a random-effects coefficient binary regression model; otherwise, a fixed-effect coefficient binary regression model was used [10].

In test accuracy studies, one of the primary causes of heterogeneity is the threshold effect, and it arises when different cut-offs or thresholds are used in different studies to define a positive (or negative) test result. The Spearman correlation coefficient between the logit of sensitivity and the logit of (1-specificity) was computed to assess the threshold effect using Review Manager 5.0 software (The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark). A strong positive
correlation would suggest a threshold effect, \( p<0.05 \) [11]. We constructed hierarchical summary receiver operating characteristic (ROC) curves to assess Sen and Spe [12]. The areas under the ROC curves (AUC) were used to analyze the diagnostic precision of CEUS for the detection of gallbladder carcinoma.

Apart from variations due to the threshold effect, there are several other factors that can result in variations in accuracy estimates amongst different test accuracy studies in a review. The presence of publication bias was visually assessed by producing a Deeks funnel plot and an asymmetry test with the STATA software. Publication bias was considered to be present if there was a nonzero slope coefficient \( (p<0.05) \) [13].

**Results**

From eighty-nine clinical studies, sixteen studies fulfilled the inclusion criteria and were included in the meta-analysis [7,14-28]. There were a total of 1673 gallbladder carcinoma cases. The search and exclusion strategy is displayed in figure 1. The characteristics of the sixteen clinical studies which were all retrospective studies are shown in table I. The quality assessment was moderate in sixteen studies according to QUADAS-2 items, and the results of the distribution of the study design are shown in figure 2.

According to the subgroups analysis, no study was excluded. The bivariate boxplot showed that no study was heterogeneous with respect to the other studies. As significant heterogeneity was found in the pooled analysis \( (I^2=50.3\%, \ p=0.067) \), Sen, Spe, PLR, and NLR were pooled by using a random effects coefficient binary regression model. The pooled weighted values were determined to be Sen: 0.92 (95%CI: 0.90), Spe 0.91 (95%CI: 0.89, 0.93), PLR 10.01 (95%CI: 7.02, 14.29), NLR 0.10 (95%CI: 0.07, 0.14), and the diagnostic odds ratio (DOR) 123.02 (95%CI: 78.40, 193.03). The forest plots from 16 studies on a per-lesion basis are shown in figure 3. ROC curves are shown in figure 4. The AUC was 0.9689 (95%CI: 0.9376, 0.9879).

A Spearman rank correlation was performed as a further test for the threshold and was determined to be 0.179 \( (p=0.561) \), which indicated that there was an absence of a notable threshold effect in the accuracy estimates among individual studies.

The results of meta-regression indicated that imaging modality and study design were not strongly associated with accuracy. The sensitivity estimates for the different subgroups are presented in table II. The sensitivity of CEUS for lesions with 1cm or small was significantly lower than that for lesions measuring larger than 1cm.
Table I. Characteristics of the clinical studies included in meta-analysis.

<table>
<thead>
<tr>
<th>Authors, year</th>
<th>Patients</th>
<th>Gender</th>
<th>Age (year)</th>
<th>Imaging modality</th>
<th>Sen</th>
<th>Spe</th>
<th>Study design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liu et al 2015</td>
<td>106</td>
<td>39 67</td>
<td>57.1+8.6</td>
<td>A</td>
<td>0.78</td>
<td>0.95</td>
<td>retrospective</td>
</tr>
<tr>
<td>Zheng et al 2013</td>
<td>86</td>
<td>37 49</td>
<td>NA</td>
<td>B</td>
<td>0.95</td>
<td>0.76</td>
<td>retrospective</td>
</tr>
<tr>
<td>Liu et al 2012</td>
<td>119</td>
<td>46 73</td>
<td>60.4+7.2</td>
<td>A</td>
<td>0.96</td>
<td>0.95</td>
<td>retrospective</td>
</tr>
<tr>
<td>Tsuji et al 2012</td>
<td>107</td>
<td>NA NA</td>
<td>63.1+6.6</td>
<td>A</td>
<td>0.94</td>
<td>0.88</td>
<td>retrospective</td>
</tr>
<tr>
<td>Xie et al 2010</td>
<td>116</td>
<td>50 66</td>
<td>54.7</td>
<td>B</td>
<td>0.85</td>
<td>0.92</td>
<td>retrospective</td>
</tr>
<tr>
<td>Hattori et al 2007</td>
<td>102</td>
<td>47 55</td>
<td>57±7</td>
<td>A</td>
<td>0.91</td>
<td>1.00</td>
<td>retrospective</td>
</tr>
<tr>
<td>Numata et al 2009</td>
<td>67</td>
<td>26 41</td>
<td>60.2+12.6</td>
<td>B</td>
<td>0.85</td>
<td>0.98</td>
<td>retrospective</td>
</tr>
<tr>
<td>Liu et al 2015</td>
<td>90</td>
<td>40 50</td>
<td>54.5</td>
<td>A</td>
<td>1.00</td>
<td>0.91</td>
<td>prospective</td>
</tr>
<tr>
<td>Hyun et al 2012</td>
<td>84</td>
<td>NA NA</td>
<td>NA</td>
<td>A</td>
<td>0.80</td>
<td>0.96</td>
<td>retrospective</td>
</tr>
<tr>
<td>Si et al 2013</td>
<td>75</td>
<td>29 46</td>
<td>53±11</td>
<td>B</td>
<td>0.85</td>
<td>1.00</td>
<td>prospective</td>
</tr>
<tr>
<td>Rifai et al 2009</td>
<td>56</td>
<td>24 32</td>
<td>56.3+7.5</td>
<td>A</td>
<td>1.00</td>
<td>0.95</td>
<td>retrospective</td>
</tr>
<tr>
<td>Kim et al 2002</td>
<td>40</td>
<td>23 38</td>
<td>54.7+10.6</td>
<td>A</td>
<td>0.92</td>
<td>0.93</td>
<td>retrospective</td>
</tr>
<tr>
<td>Inoue et al 2007</td>
<td>93</td>
<td>35 58</td>
<td>61.5+11.8</td>
<td>A</td>
<td>1.00</td>
<td>0.86</td>
<td>retrospective</td>
</tr>
<tr>
<td>Xu et al 2014</td>
<td>155</td>
<td>61 94</td>
<td>53.6</td>
<td>A</td>
<td>0.96</td>
<td>0.91</td>
<td>retrospective</td>
</tr>
<tr>
<td>Xie et al 2008</td>
<td>153</td>
<td>73 80</td>
<td>58.0</td>
<td>B</td>
<td>0.94</td>
<td>0.90</td>
<td>retrospective</td>
</tr>
<tr>
<td>Yuan et al 2007</td>
<td>184</td>
<td>72 112</td>
<td>59.3+9.5</td>
<td>A</td>
<td>0.93</td>
<td>0.79</td>
<td>prospective</td>
</tr>
</tbody>
</table>

NA, not applicable; A, Low MI with SonoVue; B, Color and Power Doppler mode

Fig 3. Forest plots of the Sensitivity, Specificity, positive likelihood ratio and negative likelihood ratio with corresponding 95% CIs for CEUS for the detection of gallbladder carcinoma.
For subgroups of imaging modality and study design, the results showed that they all had comparably high Sen estimate \((p=0.740, p=0.176, respectively)\). The mean Sen of lesions with 1 cm or smaller was 0.84 (95%CI: 81%, 88%), \(I^2=0.1\%), and for lesions larger than 1 cm, the mean Sen was 0.97 (95%CI: 94%, 99%), \(I^2=10.7\%\).

The Deeks funnel plot of standard of error by effect size for the measurements \((p=0.091)\) showed an equal distribution of studies around the middle line, indicating that publication bias was not evident (fig 4).

### Discussions

CEUS is the only imaging technique that allows continuous, dynamic, real-time observations and accurate identification of blood flow to an organ or tumor [29]. Because CEUS allows visualization of the feeding vessels to lesions and perfusion of blood, it overcomes the shortcomings of conventional ultrasonography, which cannot depict tiny blood vessels and low-velocity flows. CEUS also can distinguish tumor from sludge, being recently used to evaluate cholecystic disease [17,21]. Contrast enhancement of the polypoid changes begin in the arterial phase mainly. CEUS lead to a change of diagnosis: wall changes are found to correspond to biliary lithiasis, gallbladder polyp, part of the duodenal wall, or wall collaterals [21,23].

CEUS offers several potential advantages compared to other imaging techniques, including lower cost, non-invasive, radiation exposure, shorter scan times and most importantly the absence of nephrotoxicity. However, poor performance of ultrasound in lesion detection may be due to patient factors such as unfavourable body habits, coexistent pathology such as ascites, hernia or excessive intervening bowel gas; and graft factors such as echo reflection from the stent graft or other factors. Conversely, previous research has suggested that many of these hindrances can be minimized by the use of ultrasonic contrast. Indeed, the high false-positive rate of contrast-enhanced ultrasound compared to CT may in

![Fig 4](image_url)

**Fig 4.** Summary Receiver Operating Characteristic (SROC) curve of CEUS for the diagnosis of gallbladder carcinoma and the Deeks’ Funnel Plot Asymmetry Test for evaluating publication bias among the included studies.

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>number</th>
<th>Mean Sensitivity</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesion size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\leq1\text{cm})</td>
<td>7</td>
<td>0.84 (95%CI 81%-88%)</td>
<td></td>
</tr>
<tr>
<td>(&gt;1\text{cm})</td>
<td>9</td>
<td>0.97 (95%CI 94%-99%)</td>
<td>0.033</td>
</tr>
<tr>
<td>(\leq1\text{cm} vs. &gt;1\text{cm})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imaging modality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low MI with SonoVue (A)</td>
<td>11</td>
<td>0.94 (95%CI 91%-96%)</td>
<td></td>
</tr>
<tr>
<td>Color and Power Doppler mode (B)</td>
<td>5</td>
<td>0.89 (95%CI 87%-91%)</td>
<td></td>
</tr>
<tr>
<td>A vs. B</td>
<td></td>
<td></td>
<td>0.740</td>
</tr>
<tr>
<td>Study design</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrospective</td>
<td>13</td>
<td>0.93 (95%CI 91%-95%)</td>
<td></td>
</tr>
<tr>
<td>Prospective</td>
<td>3</td>
<td>0.90 (95%CI 88%-95%)</td>
<td>0.176</td>
</tr>
<tr>
<td>Retrospective vs. Prospective</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
fact represent its higher sensitivity in the detection of lesions [30-32].

Our meta-analysis has investigated the diagnostic performance of CEUS in gallbladder carcinoma. Analysis of 1673 lesions demonstrated consistently high pooled Sen, Spe, PLR and DOR, but it also showed a lower NLR. According to the previous research, the overall sensitivity and accuracy of CEUS was found to be slightly higher than those of CT, PET/CT and MRI [30-32]. According to these results, CEUS can be used in clinical practice as an excellent diagnostic tool for evaluation and diagnosis of gallbladder carcinoma. This finding is of great importance because of the limitations of current approaches to the identification of gallbladder carcinoma.

In our study, we found that the sensitivity of CEUS for lesions with 1cm or small was significantly lower than that for lesions measuring larger than 1cm, which is similar to the results from CT or MRI. So the diagnostic accuracy can be much improved with the technique and the resolution is improved. We also found that the sensitivity of diagnosis with low MI with SonoVue was slightly higher than that of the color and Power Doppler mode (94% vs. 89%), though the difference is not significant (p=0.740). If imaging modality is used with low MI with SonoVue by an examiner to detect gallbladder carcinoma, the relatively satisfactory results may be obtained. The difference between retrospective studies and prospective studies were not significant either. The result might be described by the fewer studies of the Doppler mode, because the methodology of CUES has changed dramatically since its introduction; the earlier high MI and Doppler mode used by Levovist are only of historical interest. Furthermore, there were not enough studies in Color and Power Doppler mode; if the studies increased, and the studies of power and color Doppler modes that were used with Levovist were excluded, diagnostic sensitivity of CEUS using low MI and SonoVue, might have improved, and the result may have revealed the real clinical value.

This meta-analysis was based on a rigorous literature search, which resulted in the inclusion of sixteen articles, and a validated appraisal tool was used to determine the risk of bias of these included studies. Still, there are some limitations. With the statistical package used in the present study, we were able to calculate confidence and credible intervals for the overall sensitivity and specificity, but not for the likelihood ratios. No method was found to calculate CIs around the likelihood ratios that are derived from the overall estimates of sensitivity and specificity. Other limitations are that some studies were specifically excluded because of incomplete data reporting (unable to construct a 2×2 table). There were also recurrent sources of bias on three of the four items of the QUADAS tool, which shows poor reporting of participants’ characteristics and a study design for the included studies. Funnel plot or Egger testing that were used in publication bias were controversial because there were many reasons that could cause the asymmetry, such as opportunity, heterogeneity, effect of choice, the choice of measurement precision, etc [33,34]. The experience of clinical epidemiology was of little help to judge result; subjective visual judgment was the main method. The following factors may lead to inefficiency of funnel plot or egger testing: 1) many factors lead to asymmetric funnel plot, which may be more heterogeneity rather than publish bias. In meta-analysis, it is difficult to ensure that various studies included in the analysis use the same method and the same quality. There is no guarantee that the patients owned the same disease severity, even the same personality traits, such as culture, geographical and racial differences were difficult to control. So we have no reason to expect the scatter diagram to be a symmetric shape of a funnel; 2) the effect size will offer great influence on graphic shapes; 3) if there were some potential relations between the size of sample size and effect size, the wrong conclusions could be drawn. Finally, while the two raters were fairly concordant when evaluating the risk of bias (κ>0.50), one item was less concordant. However, agreement was easily obtained between the two raters and the initial disagreements were mostly related to adherence to the defined standard.

Conclusions

Our meta-analysis demonstrated that CEUS is a liable, non-invasive, and non-radiative imaging modality with a high sensitivity and specificity for the detection of gallbladder carcinoma. Nonetheless, as a relatively new technique, considering the notable heterogeneity and the existing inherent limitations, it should be applied cautiously, and large scale, well-designed trials are necessary to assess its clinical value.

Conflicts of interest: none

Reference