

Meta-analysis: Contrast-enhanced Ultrasound Versus Conventional Ultrasound for Differentiation of Benign and Malignant Breast Lesions

Qian Li^{a¶}, Min Hu^{b¶}, Zhikui Chen^{c¶}, Changtian Li^d, Xi Zhang^e, Yiqing Song^f, Feixiang Xiang^g *

^aDepartment of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

^bDepartment of Cardiovascular Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, People's Republic of China.

^cDepartment of Ultrasound, Union hospital, Fujian medical university, Fuzhou, Fujian, China

^dDepartment of Ultrasound, The Southern Building, Chinese PLA General Hospital, Beijing, China.

^eClinical Research Unit, Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China.

^fDepartment of Epidemiology, Indiana University Richard M. Fairbanks School of Public Health, Indianapolis, United States of America

^gDepartment of Ultrasound, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

*Corresponding author

Email: xiangfx@hotmail.com

This is the author's manuscript of the article published in final edited form as:

Li, Q., Hu, M., Chen, Z., Li, C., Zhang, X., Song, Y., & Xiang, F. (2018). Meta-Analysis: Contrast-Enhanced Ultrasound Versus Conventional Ultrasound for Differentiation of Benign and Malignant Breast Lesions. *Ultrasound in Medicine & Biology*, 44(5), 919–929. <https://doi.org/10.1016/j.ultrasmedbio.2018.01.022>

Abstract-This meta-analysis aimed to compare the diagnostic performance of contrast-enhanced ultrasound (CEUS), conventional ultrasound (US) combined with CEUS (US+CEUS), and US for distinguishing breast lesions. From thorough literature research, studies that compared the diagnostic performance of CEUS vs. US or US+CEUS vs. US, using pathology results as the gold standard, were included. A total of 10 studies were included, of which 9 compared the diagnostic performance of CEUS and US, and 5 studies compared US+CEUS and US. In those comparing CEUS vs. US, the pooled sensitivity was 0.93 (95%CI: 0.91-0.95) vs. 0.87 (95%CI: 0.85-0.90), and pooled specificity was 0.86 (95%CI: 0.84-0.88) vs. 0.72 (95%CI: 0.69-0.75). In studies comparing US+CEUS vs. US, the pooled sensitivity was 0.94 (95%CI: 0.92-0.96) vs. 0.87 (95%CI: 0.84-0.90), and pooled specificity was 0.86 (95%CI: 0.82-0.89) vs. 0.80 (95%CI: 0.76-0.84). In terms of diagnosing breast malignancy, areas under the SROC curve (AUC) of both CEUS (P=0.003) and US+CEUS (P=0.000) were statistically higher than that of US. Both CEUS alone and US+CEUS had better diagnostic performance than US in differentiation of breast lesions, and US+CEUS also had low negative likelihood ratio.

Key Words: Breast, Ultrasonography, Contrast, Diagnosis, Meta-analysis

Introduction

Globally, breast cancer is the most common cancer and the leading cause of cancer death among women (Torre et al. 2015). High-frequency ultrasonography has become the first-line imaging modality in evaluation of breast lesions due to its widespread availability, non-invasiveness, and low expense. However, conventional ultrasound (US) faces some limitations in differentiating benignity from malignancy because of overlapping sonographic findings in some cases (Zhi et al. 2007). Unlike conventional US, the newly emerging contrast-enhanced ultrasound (CEUS) helps evaluate blood distribution and perfusion of tumors, thus offering more valuable information for lesion differentiation (Harvey et al. 2015; Ishii et al. 2017; Lekht et al. 2016).

However, the capability of CEUS to accurately diagnose breast cancer remains unclear. A meta-analysis of 16 studies found that the pooled sensitivity and specificity of CEUS alone in diagnosing breast cancers were 0.86 and 0.79 (Hu et al. 2014), which were similar to the sensitivity (0.82-0.95) and specificity (0.71-0.79) of conventional US reported in several studies (Du et al. 2012; Liu et al. 2008; Xiao et al. 2016). This difference was ascribed to CEUS' capability to delineate the morphological features of breast masses, which conventional US does not possess. Therefore, in order to make full use of the sonographic information offered by each technique, conventional US and CEUS (US+CEUS) were combined. While several studies found improved sensitivity (US+CEUS: 0.88-0.97 vs. US: 0.82-0.89) and

specificity (US+CEUS: 0.82-0.93 vs. US: 0.78-0.79) (Du et al. 2012; Wang et al. 2011; Xiao et al. 2016), no improvement was found in other studies (Fujimitsu et al. 2016; Sorelli et al. 2010).

Until now, no meta-analysis has compared the diagnostic performance of CEUS and US or US+CEUS and US in differentiating breast cancers. Here we systematically reviewed the literature via a meta-analysis to compare the diagnostic performance of these ultrasound techniques on benign and malignant breast lesions, using pathological results as the reference standard.

Methods

Our meta-analysis was conducted in accordance with the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) (Moher 2009). We included studies of patients suspected of having one breast mass or more, using US and CEUS, or US and US+CEUS as diagnostic methods, histopathological or cytological results for comparison, and reporting true positive (TP), false positive (FP), true negative (TN), and false negative (FN), and the study type was diagnostic test.

Search Strategy

We searched online all published studies without language restrictions from the earliest available date of indexing to August 31, 2016 in PubMed, EMBASE, and Cochrane Library databases. We also searched the most comprehensive Chinese academic databases in medicine: China National Knowledge Infrastructure, Chinese Biomedical Literature Database, and Wanfang Database. The references of retrieved

articles were also hand-searched. The search strategy included the following terms: (“contrast enhanced ultrasound” OR “contrast enhanced ultrasonography” OR “CEUS”) in combination with (“breast”) and (“ultrasound” OR “ultrasonography” OR “sonography”). The search strategy for Chinese papers was similar to that used for English papers (S1 File).

Study Selection

All published studies that compared the diagnostic accuracy of CEUS, US+CEUS, and US for breast lesions were identified. Study selection was performed independently by two researchers. If disagreements occurred, a third reviewer made the adjudication. First, the titles and abstracts were screened to determine the potential usefulness of the articles, followed by full-text screening according to the following criteria (1) patients: suspect of having breast mass, (2) studies did obtain informed consent from each study participant, and each study was approved by an ethics committee or institutional review board, (3) index tests: both US and CEUS, or both US and US+CEUS were used for diagnostic purposes, (4) studies compared the performance of CEUS and US or US+CEUS and US for differentiating benign from malignant breast masses, (5) studies of harmonic-mode CEUS, (6) reference standard was either histopathology or cytology, and (7) TP results and FN results or TN results and FP results were available or could be derived adequately. Exclusion criteria were (1) case reports or case series, review articles, letters, comments, (2) studies of contrast-enhanced power or color Doppler sonography, (3) duplicate publications in different databases and studies using the same

study population from the same institution, (4) fewer than 15 cases confirmed by reference standard, and (5) postsurgical studies.

Data Extraction

Two authors independently extracted the data from eligible studies, and discrepancies were resolved by discussion. Adjudication by a third investigator was performed when disagreements occurred. Extracted information included (1) first author name, (2) year of publication, (3) age, (4) number of patients, (5) number of masses, (6) total number of malignant masses, (7) mass long axis, and (8) reference standard. The diagnostic accuracy data on each index test and number of TP, TN, FP, and FN findings for each index test were recorded or calculated. Sensitivity, specificity, positive likelihood ratio (LR+), and negative likelihood ratio (LR-) were extracted or calculated as follows: $\text{sensitivity} = \text{TP}/(\text{TP} + \text{FN})$, $\text{specificity} = \text{TN}/(\text{TN} + \text{FP})$, $\text{LR+} = \text{sensitivity}/(1 - \text{specificity})$, and $\text{LR-} = (1 - \text{sensitivity})/\text{specificity}$. If more than one CEUS criterion was used in one individual literature, the data of the one with the highest Youden index or area under the curve were extracted or calculated.

To ensure the consistency of subjects and methods in the included studies, we extracted, pooled, and compared the diagnostic accuracy of US and CEUS from the literature that compared these two techniques (Group 1), and studies comparing US and US+CEUS were classified into another group (Group 2).

Quality Assessment

All included studies were assessed for methodological quality by two authors

independently using the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) tool (Whiting 2011). If the two readers disagreed, a third reader adjudicated. None of the readers was involved in any of the included studies. The QUADAS-2 checklist consists of four domains: patient selection, index test, reference standard, and flow/timing. Based on several questions, each of the four domains was assessed for risk of bias, but only the first three domains were assessed for applicability concerns.

Statistical Analysis

Sensitivity, specificity, LR+, and LR- were extracted or calculated from individual studies and then pooled to assess diagnostic accuracy. Summary receiver operating characteristic (SROC) curves were constructed, also to examine diagnostic accuracy. Between-study heterogeneity was tested using inconsistency index (I^2) statistics, and I^2 values greater than 50% were considered to indicate substantial heterogeneity (Higgins 2003). A random-effects model was applied in the analysis for the heterogenic data, and the fixed-effects model was used otherwise (DerSimonian and Laird 1986).

Likelihood ratio (LR) can be interpreted as follows: a LR of 0 excludes disease, a LR of infinity (∞) excludes normality, and a LR of 1 means no change in likelihood of disease. For the diagnostic information to have high probability of altering clinical management, a LR greater than 10 or less than 0.1 would be required for a positive or negative test result, respectively. Moderate informational value can be achieved with LRs of 5-10 and 0.1-0.2, and LRs of 2.0-5.0 and 0.2-0.5 indicate very little informational value (Sadigh et al. 2012).

A Deek's funnel plot was used to assess the publication bias of all included studies. Z-test was used to test differences of area under SROC curve (AUC) between CEUS and US, or between CEUS+US and US.

A sensitivity analysis was performed to investigate the influence of the publication language on the diagnostic performance of different ultrasound techniques.

In CEUS studies, Subgroup analysis was performed on the dosage of SonVue, which was used as the contrast in 8 of these 9 studies. Meanwhile, Subgroup analysis was performed on diagnostic criteria for malignancy in US studies.

The meta-analysis was performed using Meta Disc version 1.4 (Meta-DiSc, Javier.zamora), and Stata 14 was used to analyze publication bias. MedCalc Statistical Software version 13.0.2 (MedCalc Software bvba, Ostend, Belgium; <http://www.medcalc.org>) was used to compare SROCs. All statistical tests were two-sided, and statistical significance was defined as $P < 0.05$.

Results

Search Results and Selected Articles

A total of 1359 articles were identified from the database. After primary title and abstract screening, 54 studies were identified for further full-text evaluation, finally leaving 10 qualified studies in this meta-analysis (**Fig. 1**).

The 10 included studies were classified into two groups according to the ultrasound techniques evaluated in each study. Of the 9 studies in Group 1 that compared the diagnostic accuracy of CEUS vs. US, 7 were conducted in China and the other 2 in Italy

and Japan, respectively. Of the 5 studies in Group 2 that compared the diagnostic accuracy of US+CEUS vs. US, all were conducted in China. The diagnostic accuracy of CEUS vs. US and US+CEUS vs. US were both compared in 4 studies.

In Group 1, with the exception of one male, all patients were female (Ricci et al. 2007) (age range 19-88 years), including 1545 patients and 1609 breast masses, of which 751 masses (751/1609, 46.7%) were malignant. In Group 2, all patients were female (age range 19-88), including 924 patients and 959 breast masses, of which 505 masses (505/959, 52.6%) were malignant.

Quality Assessment

The characteristics of all 10 studies are shown in Table 1. The methodological quality of the included papers was assessed using the QUADAS-2 tool (**Fig. 2**). In the Group 1 studies (CEUS vs. US), one study was scored high risk because the threshold for malignancy of US and CEUS was not pre-specified (An et al. 2014), while 2 studies were scored as having unclear risk because it was not clear whether the US threshold was pre-specified (Miyamoto et al. 2014; Ricci et al. 2007). In Group 2 studies (US+CEUS vs. US), one study (An et al. 2014) was scored high risk and another study (Le et al. 2012) unclear risk, for the above reasons.

Analyses of studies in Group 1 (CEUS vs. US)

Diagnostic Accuracy of CEUS for Malignant Breast Lesions

To diagnose malignant breast lesions, the sensitivity and specificity of CEUS in individual studies ranged from 0.76-1.00 and 0.82-0.97 (respectively), and the pooled

sensitivity and specificity of CEUS were 0.93 (95% CI: 0.91-0.95) and 0.86 (95% CI: 0.84-0.88). The included studies were statistically heterogeneous in the estimate of sensitivity ($I^2=61.2\%$) but not in the estimate of specificity ($I^2= 22.4\%$) (**Fig. 3**). The pooled LR+ and LR- were 6.48 (95%CI: 5.55-.57) and 0.10 (95%CI: 0.06-0.16), respectively.

Diagnostic Accuracy of US for Malignant Breast Lesions

In the 9 studies that compared the performance of CEUS and US to diagnose malignant breast lesions, the sensitivity and specificity of the US in individual studies ranged from 0.69-0.95 and 0.58-0.92 (respectively), and the pooled sensitivity and specificity of US were 0.87 (95% CI: 0.85-0.90) and 0.72 (95% CI: 0.69-0.75). The included studies were statistically heterogeneous in the estimate of sensitivity ($I^2=78.6\%$) and specificity ($I^2=84.6\%$) (**Fig. 3**). The pooled LR+ and LR- were 3.50 (95%CI: 3.59-4.73) and 0.20 (95%CI: 0.13-0.31), respectively.

Publication Bias

The Deek's funnel plot asymmetry test showed no publication bias among CEUS ($P=0.74$) and US ($P=0.56$) studies.

Comparison of the Accuracy of CEUS and US for Detection of Malignant Breast Lesions

The area under the SROC curve (AUC) of CEUS (AUC=0.954, SE=0.008) of the 9 studies that compared the diagnostic performance of CEUS and US was significantly higher than that of US (AUC=0.884, SE=0.022) ($Z=3.012$, $P=0.003$) (**Fig. 4**).

Subgroup Analysis

In subgroup analysis of CEUS studies (**Table 2**), the heterogeneity of sensitivity decreased in the group with dosage ≤ 2.4 ml but increased in the group with dosage > 2.4 ml, and the heterogeneity of specificity increased in both groups. Meanwhile, there was no significant difference in sensitivity or specificity between these two subgroups ($P=0.368$). Subgroup analysis was not performed on paper quality (assessed by QUADAS-2) or study design, due to limited sample size in subgroups.

In subgroup analysis of US studies (**Table 2**), heterogeneity was still observed in sensitivity and specificity even when lesions were stratified by diagnostic criteria for malignancy, and there was no significant difference in sensitivity or specificity between studies using BIRADS $\geq 4b$ as diagnostic criteria and studies that did not use it ($P=0.147$). Also, subgroup analysis was not performed on paper quality (assessed by QUADAS-2) or study design, due to limited sample size in subgroups.

Sensitivity Analysis

A sensitivity analysis was performed to investigate the influence of the publication language on the diagnostic performance of different ultrasound techniques. After excluding the studies published in Chinese, the sensitivity and specificity of CEUS in studies published in English were 0.93 (95%CI:0.91-0.95) and 0.88 (95%CI: 0.85-0.91), and the sensitivity and specificity of US were 0.87 (95%CI: 0.83-0.90) and 0.72 (95%CI: 0.69-0.74), both of which were similar to figures in overall studies, i.e., CEUS: 0.93 (95%CI: 0.91-0.95) and 0.86 (95%CI: 0.84-0.88), US: 0.87 (95%CI: 0.85-0.90) and

0.72 (95%CI: 0.69-0.75).

Analyses of Studies in Group 2 (US+CEUS vs. US)

Diagnostic Accuracy of US+CEUS for Malignant Breast Lesions

The sensitivity and specificity of US+CEUS to diagnose malignant breast lesions in individual studies ranged from 0.85-0.98 and 0.82-0.90, respectively. The pooled sensitivity and specificity of US+CEUS were 0.94 (95%CI: 0.92-0.96) and 0.86 (95%CI: 0.82-0.89), respectively. The included studies were statistically heterogeneous in their estimate of sensitivity ($I^2=80.5\%$) but not of specificity ($I^2=40.0\%$) (**Fig. 5**). The pooled LR+ and LR- were 6.43 (95%CI: 5.14-8.04) and 0.07 (95%CI: 0.03-0.18), respectively.

Diagnostic Accuracy of US for Malignant Breast Lesions

In the 5 studies in Group 2, the sensitivity and specificity of US to diagnose malignant breast lesions in individual studies ranged from 0.78-0.95 and 0.75-0.88, respectively. The pooled sensitivity and specificity of US were 0.87 (95%CI: 0.84-0.90) and 0.80 (95%CI: 0.76-0.84), respectively. The included studies were statistically heterogeneous in the estimate of sensitivity ($I^2= 64.0\%$) but not of specificity ($I^2= 41.1\%$) (**Fig. 5**). The pooled LR+ and LR- were 4.39 (95%CI: 3.63-5.29) and 0.17(95%CI: 0.11-0.25), respectively.

Publication Bias

The Deek's funnel plot asymmetry test showed no publication bias among studies for US+CEUS ($P=0.82$) and US ($P=0.62$).

Comparison of the accuracy of US and US+CEUS for malignant breast lesions

The AUC of SROC of US+CEUS (AUC=0.965, SE=0.009) was significantly higher than that of US (Group 2) (AUC=0.911, SE=0.011) ($Z=3.826$, $P=0.000$) (Fig. 6).

Sensitivity Analysis

After excluding studies published in Chinese, the sensitivity and specificity of US+CEUS in studies published in English were 0.96 (95%CI: 0.94-0.98) and 0.84 (95%CI: 0.80-0.88), and the sensitivity and specificity of US were 0.89 (95%CI: 0.85-0.92) and 0.77 (95%CI: 0.72-0.82), both of which were similar to findings in overall studies (US+CEUS: 0.94 (95%CI: 0.92-0.96) and 0.86 (95%CI: 0.82-0.89), US: 0.87 (95%CI: 0.84-0.90) and 0.80 (95%CI: 0.76-0.84)).

Discussion

Our meta-analysis compared the diagnostic performance of CEUS, US, and US+CEUS for differentiating benign from malignant breast lesions, using pathology as the reference standard. However, the diagnostic accuracy of these ultrasound techniques varies among different studies due to variation in examiners, types of CEUS contrast, analyzing software, etc. (Du et al. 2012; Hu et al. 2014; Ma et al. 2015; Ricci et al. 2007). To minimize data heterogeneity due to the study population, we reclassified the studies into two groups in this meta-analysis: Group 1, which compared the diagnostic accuracy of CEUS vs. US; and Group 2, which compared US+CEUS vs.

US, in order to compare the diagnostic performance in the same study population of each group (Alrajab et al. 2013; Schwab et al. 2016).

The acquisition modes during CEUS scanning, including power Doppler, color Doppler, and harmonic modes, might also have introduced variation among studies. In this study, we included only studies of CEUS that were performed with the harmonic mode, for the following considerations. Compared to power or color Doppler mode, the harmonic mode has fewer limitations in the frame rate and motion artifacts (Ma et al. 2016). And additionally, harmonic mode detects signals using a low mechanical index, which causes minimal bubble destruction and allows continuous real-time assessment of the microvascularization (Cosgrove and Blomley 2004).

Our data showed that most of the included papers published in English were actually performed in China (6/8, 75%), which was similar to the findings of two other meta-analyses assessing the diagnostic performance of CEUS on breast lesions, with 10/15 (Ma et al. 2016) and 4/6 (Hu et al. 2014) performed in China, respectively. The abundance of Chinese studies in this area may be ascribed to the early approval of SonVue in the Chinese market and the popularity of ultrasound exams in Chinese clinical practice (Xu and Lu 2010). In addition, we performed searches in Chinese databases, which added one study to Group 1 and two studies to Group 2. As our sensitivity analysis showed no change in diagnostic performance when studies published in Chinese were excluded, all the studies published in Chinese were also included for final analysis.

As the first-line breast-imaging procedure, conventional US has shown high accuracy in distinguishing between benign and malignant breast lesions (Costantini et al. 2006), which is based on its capability to evaluate lesion features including morphology, echotexture, and vascular distribution. Using conventional high-frequency US, morphological and echotexture features of malignant breast lesions have been fully studied, including lesion shape, orientation, margins, boundary, echo pattern, and posterior acoustic features (Costantini et al. 2006). In addition, some ultrasound technologies such as color and power Doppler imaging have been widely used to evaluate blood distribution in lesions. It has been demonstrated that angiogenesis plays an important role in the development of breast cancer in terms of tumor growth and metastasis (Drudi et al. 2012), including most of the advanced breast cancers, and some others at early stages, such as atypical hyperplasia or lobular *carcinoma in situ* (Drudi et al. 2012). Researchers have found that internal and surrounding vascularity on ultrasound are characteristic of malignancy (Ferrara et al. 2016; Sehgal et al. 2000), and vascular density is associated with tumor aggressiveness (Balleyguier et al. 2009). Together these sonographic features of conventional US offer relatively high accuracy in distinguishing between benign and malignant breast lesions.

However, the evaluation of vascular distribution in tumors using power and color Doppler are not acceptable due to their low sensitivity in detecting the small vessels and slow blood flow associated with tumor neovascularity (Ferrara et al. 2016; Sehgal et al. 2000). Fortunately, ultrasound contrast agents are microbubbles 1-10 microns in

size (equal to or smaller than red blood cells), enabling the visualization of both the macrovasculature and microvasculature associated with tumors. Our results showed that CEUS alone had better diagnostic performance than US alone (AUC: 0.954 vs. 0.884), which suggested that CEUS was helpful in differentiating malignant from benign breast lesions. Similarly, it has been reported that, compared to color Doppler in conventional US (sensitivity 83.8-95.3%, specificity 57.7-77.7%), CEUS alone has superior diagnostic performance in differentiating between benign and malignant breast lesions (sensitivity 91.4-95.5%, specificity 81.6-88.3%) (Liu et al. 2008; Miyamoto et al. 2014; Xiao et al. 2016).

It should be noted that the diagnostic performance of CEUS alone may be overestimated, because clinically, a 2D ultrasound screening is usually performed before CEUS, and the 2D sonographic information may inadvertently affect the evaluation of the diagnostic performance of CEUS alone.

In addition to the advantages of CEUS compared to conventional US, CEUS examination alone has several limitations, including the inability to delineate the morphological characteristics of lesions and the variation in vascular distribution patterns in tumors with different histopathologic types (Wang et al. 2011). For example, invasive ductal carcinoma, which accounts for the majority of breast malignancies, showed heterogeneous enhancement on CEUS. Conversely, most medullary carcinomas showed homogenous enhancement (Wang et al. 2011). However, because invasive ductal carcinomas have irregular shape and poorly defined margins, while

medullary carcinomas have regular shapes and well-defined margins, CEUS alone would not be a sufficiently sensitive tool with which to differentiate special types of cancers. Therefore it is reasonable to combine US with CEUS, which makes full use of the advantages of both techniques. Our pooled data showed that the sensitivity and specificity of CEUS+US were 0.94 (95% CI: 0.92-0.96) and 0.86 (95% CI: 0.82-0.89), respectively, and the AUC of SROC of CEUS+US was higher than that of US (0.965 vs. 0.911), which is consistent with the findings of several other researchers, underlining the advantages of CEUS+US over US alone (Du et al. 2012; Liu et al. 2008; Xiao et al. 2016).

Generally, a helpful diagnostic test should have a high positive LR (>5 : good at ruling in a disease) and low negative LR (<0.2 : good at ruling out a disease) (Cronin et al. 2008). Compared to US, CEUS and US+CEUS are better at confirming malignancy due to the positive LR+, and are also helpful in ruling out malignancy due to lower LR-. Additionally, as US+CEUS with LR- <0.1 is highly capable of excluding all malignancy, a negative US+CEUS usually indicates that no further diagnostic tests are required.

Heterogeneity

In Group 1, heterogeneity was found in the sensitivity and specificity of US, and in the sensitivity of CEUS, which persisted even after subgroup analysis. In Group 2, heterogeneity was found in the sensitivity of both US and US+CEUS, but subgroup analysis was not conducted because of the limited number of studies included.

For US, our results showed that neither use of BIRADS $\geq 4b$ as the diagnostic

criterion for malignancy, paper quality assessed by QUADAS-2, nor prospective/retrospective study design were sources of heterogeneity. Heterogeneity may have arisen from the inclusion criteria, ultrasound devices, and probe frequencies. The variations among studies in the stage of included lesions may introduce differences in malignant frequency in study populations, affecting the diagnostic performance of an ultrasound technique (Du et al. 2012; Liu et al. 2008; Xiao et al. 2014; Xiao et al. 2016; Zhang et al. 2014). In the studies comprising this meta-analysis, a total of 8 ultrasound systems (Esatune (Ricci et al. 2007), Esaote (Zhao et al. 2010), HDI 5000 (Liu et al. 2008), Philips IU22 (An et al. 2014; Liu et al. 2008; Xiao et al. 2014), AplioXG (Miyamoto et al. 2014), Prosound α 10 (Miyamoto et al. 2014), LOGIQ E9 and LOGIQ 9 (Zhang et al. 2014) were used, with the probe frequencies ranging from 5MHz to 15MHz, either of which might cause the heterogeneity among studies.

As to CEUS, our studies showed that neither the dosage of SonVue, paper quality assessed by QUADAS-2, nor prospective/retrospective design were sources of heterogeneity. Potential heterogeneity sources may include differences in contrast techniques, software for CEUS imaging processing, or criteria for diagnosing malignancy. Here to, different manufactures use different contrast techniques, such as Contrast-Tuned Imaging (CnTI) of Esaote (Du et al. 2012; Le et al. 2012; Ricci et al. 2007; Zhao et al. 2010), Coded Angio Harmonic (CAH) of GE (Miyamoto et al. 2014; Zhang et al. 2014), and Pulse Inversion Harmonic Imaging (PIHI) of Philips (An et al. 2014; Liu et al. 2008; Xiao et al. 2014; Xiao et al. 2016), etc. Furthermore, different

quantitative analysis software was used in some studies, such as Q-lab (An et al. 2014; Liu et al. 2008) and Qontrast (Du et al. 2012; Le et al. 2012; Ricci et al. 2007; Zhao et al. 2010). The most common criteria for diagnosing malignancy in CEUS examinations were (1) peripheral enhancement, based on more peripheral micro-vessel distribution in the tumor's angiogenesis, and (2) the characteristic enhancement pattern with early, intense wash-in and fast wash-out phases, associated with a tumor's arteriovenous shunts (Drudi et al. 2012). In addition, other criteria for malignancy were used in some studies, such as strengthening of enhancement (Zhang et al. 2014), scoring system (Du et al. 2012; Xiao et al. 2014), and inhomogeneous enhancement (An et al. 2014)

Limitations

There were several limitations to our study. First, only 10 eligible studies were included in two groups (9 and 5 in Group 1 and 2, respectively and 4 in both), and this small sample may limit the power of the data analysis or the generalizability of the study findings. Second, most eligible studies were conducted in Asia (8 of 9 CEUS vs. US studies and 5 of 5 US+CEUS vs. US studies), and this uneven geographical distribution may cause variations because of the health-care gap between regions. Also, the differences in cancer pathology between ethnic groups may be another source of variation. Third, there is a limitation in our analysis. The diagnostic performances of CEUS and US+CEUS were not compared due to the limited number of studies included, and in addition, no subgroup analysis was conducted in US+CEUS vs. US studies, due to the limited numbers of studies in subgroups. Finally, study quality assessment

showed one study in Group 1 (1/9) and another in group 2 (1/5) had high risk, because the threshold for malignancy of US and CEUS was not pre-specified, but they had limited impact on the meta-analysis result due to the number.

Conclusions

Our meta-analysis shows that both CEUS alone and US+CEUS have better AUC of SROC than conventional US for distinguishing between benign and malignant breast lesions, and both have excellent sensitivity. Meanwhile, US+CEUS has a low negative likelihood ratio, which recommends its use for ruling out breast malignancy. However, our meta-analysis is based on a small number of studies, many of which were performed in a single geographic region. Further studies are therefore needed before generalization of this conclusion.

Acknowledgments

This work was supported by “the Fundamental Research Funds for the Central Universities” in China under award numbers 2017KFYXJJ255. The authors wish to acknowledge Jian Wu and Min Wu for manuscript editing.

Table 1. Characteristics of included studies

Study	Group	Year	Country	Study design	No. of patients	Patient age, min-max (mean), years	No. of masses	Malignant, no. (% of all masses)	Mass long axis, min-max(mean), mm	Reference standard
Ricci et al. (2007)	1	2007	Italy	Pro	48	40-81(58)	50	26(52.0)	ND	Histo
Liu et al. (2008)	1&2	2008	China	Pro	103	19-86(44)	103	43(41.7)	5-76	Histo
Zhao et al. (2010)	1	2010	China	Pro	71	30-78(/)	76	45(63.4)	ND	Histo
Le et al. (2012)	2	2012	China	Retro	197	13-81(48)	198	97(49.2)	ND	Histo
Du et al. (2012)	1&2	2012	China	Pro	61	27-71(48)	61	33(54.1)	8-27	Histo
An et al. (2014)	1&2	2014	China	Retro	73	19-68(44)	73	41(56.2)	ND	Histo
Xiao et al 1. (2014)	1	2014	China	Pro	475	16-84(43)	498	207(54.6)	3-49	Histo
Miyamoto et al. (2014)	1	2014	Japan	Pro	117	22-79(46)	117	35(29.9)	4-34	Histo/cyto
Zhang et al. (2014)	1	2014	China	Pro	107	31-71(ND)	107	30(28.0)	3-10	Histo
Xiao et al 2. (2016)	1&2	2016	China	Pro	490	21-88(46)	524	291(55.5)	2.8-56	Histo

Group 1: studies compared CEUS vs. US

Group 2: studies compared US+CEUS vs. US

Pro: prospective design

Retro: retrospective design

ND: not determined

Histo: histopathology

Cyto: cytopathology

Table 2. Subgroup analysis for the diagnostic performance of CEUS and US to differentiate benign and malignant breast lesions in studies in Group 1.

Ultrasound techniques	Parameter	Studies, n	Sensitivity (95%CI)	I ² , %	Specificity (95%CI)	I ² , %	P for interaction	
CEUS	Overall	9	0.93 (0.91-0.95)	61.2	0.86 (0.84-0.88)	22.4		
	Dosage of SonVue*							
	≤2.4 ml	3	0.91 (0.84-0.95)	5.6	0.89 (0.83-0.93)	40.6	0.368	
	>2.4 ml	5	0.94 (0.91-0.95)	76.3	0.85 (0.82-0.88)	28.5		
	QUADAS							
	Low risk	8	0.93 (0.91-0.95)	65.4	0.86 (0.84-0.88)	31.7	/	
	Unclear risk	0	/	/	/	/	/	
	High risk	1	/	/	/	/	/	
	Study design							
	Pro	8	0.93 (0.91-0.95)	65.4	0.88 (0.85-0.91)	31.7	/	

	Retro	1	/	/	/	/	
US	Overall	9	0.87 (0.85-0.90)	78.6	0.72 (0.69-0.75)	84.6	
	Diagnostic Criteria for malignancy						
	BIRADS \geq 4b	3	0.90 (0.87-0.92)	86.0	0.76 (0.73-0.80)	88.8	0.147
	Others [#]	6	0.83 (0.78-0.87)	66.4	0.66 (0.61-0.70)	75.5	
	QUADAS						
	Low risk	6	0.89 (0.86-0.91)	82.8	0.77 (0.73-0.80)	72.8	/
	Unclear risk	2	/	/	/	/	
	High risk	1	/	/	/	/	
	Study design						
	Pro	8	0.87 (0.83-0.90)	81.1	0.72 (0.69-0.74)	85.8	/
	Retro	1	/	/	/	/	

*One study using Sonazoid (Miyamoto et al. 2014) as the contrast was excluded from the subgroup

analysis stratified by dosage of SonVue

“Other” subgroup includes one study regarding BIRADS > 3 as malignancy, and five studies, in

which BIRADS was not used as the diagnostic criteria

I²: inconsistency index

CEUS: contrast-enhanced ultrasound

US: conventional ultrasound

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

Pro: prospective design

Retro: retrospective design

BIRADS: Breast Imaging Reporting Data System

References

- Alrajab S, Youssef AM, Akkus NI, Caldito G. Pleural ultrasonography versus chest radiography for the diagnosis of pneumothorax: review of the literature and meta-analysis. *Crit Care Critical Care* 2013;17:R208–1.
- An S, Liu J, Chen Q, Man L, Guo Z. Conventional ultrasound combined with contrast-enhanced ultrasound in diagnosis of benign and malignant breast tumors. *Chin J Med Image* 2014;29:1774–1777.
- Balleyguier C, Opolon P, Mathieu MC, Athanasiou A, Garbay JR, Delalogue S, Dromain C. New potential and applications of contrast-enhanced ultrasound of the breast: Own investigations and review of the literature. *Eur J Radiol*, 2009;69:14–23.
- Cosgrove D, Blomley M. Liver tumors. *Abdom Imaging* 2004;29:446–454.
- Costantini M, Belli P, Lombardi R, Franceschini G, Mule A, Bonomo L. Characterization of Solid Breast Masses Use of the Sonographic Breast Imaging Reporting and Data System Lexicon. *J Ultrasound Med* 2006;25:649–659.
- Cronin P, Dwamena BA, Kelly AM, Bernstein SJ, Carlos RC. Solitary pulmonary nodules and masses: a meta-analysis of the diagnostic utility of alternative imaging tests. *Eur Radiol* 2008;18:1840–1856.
- DerSimonian R and Laird N. Meta-analysis in clinical trials. *Control Clin Trials*

1986;7: 177-88

Drudi F, Cantisani V, Gneccchi M, Malpassini F, Di Leo N, de Felice C. Contrast-Enhanced Ultrasound Examination of the Breast: A Literature Review. *Ultraschall Med* 2012;33:E1–E7.

Du J, Wang L, Wan C-F, Hua J, Fang H, Chen J, Li F-H. Differentiating benign from malignant solid breast lesions: Combined utility of conventional ultrasound and contrast-enhanced ultrasound in comparison with magnetic resonance imaging. *European Journal of Radiology* 2012;81:3890–3899.

Ferrara KW, Merritt CRB, Burns PN, Foster S, Mattrey RF, Wickline SA. Evaluation of Tumor Angiogenesis with US: Imaging, Doppler, and Contrast Agents. *Acad Radiol* 2016;7:824–839.

Fujimitsu R, Shimakura M, Urakawa H, Morita A, Shinagawa Y, Sakamoto K, Yoshimitsu K. Homogeneously enhancing breast lesions on contrast enhanced US: differential diagnosis by conventional and contrast enhanced US findings. *Jpn J Radiol* 2016;34:508–514.

Harvey CJ, Alsafi A, Kuzmich S, Ngo A, Papadopoulou I, Lakhani A, Berkowitz Y, Moser S, Sidhu PS, Cosgrove DO. Role of US Contrast Agents in the Assessment of Indeterminate Solid and Cystic Lesions in Native and Transplant Kidneys.

- Radiographics 2015;35:1419–1430.
- Higgins JPT. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–560.
- Hu Q, Wang XY, Zhu SY, Kang LK, Xiao YJ, Zheng HY. Meta-analysis of contrast-enhanced ultrasound for the differentiation of benign and malignant breast lesions. *Acta Radiologica* 2014;56:25–33.
- Ishii T, Numata K, Hao Y, Doba N, Hara K, Kondo M, Tanaka K, Maeda S. Evaluation of hepatocellular carcinoma tumor vascularity using contrast-enhanced ultrasonography as a predictor for local recurrence following radiofrequency ablation. *Eur J Radiol* 2017;89:234–241.
- Le J, Chang C, Chen M, Fan Y, Zhou S, Shen J. Comparison of conventional ultrasound combined with contrast-enhanced ultrasound and conventional ultrasound in differential diagnosis of breast lesion. *Shanghai Med Imaging* 2012;21:182–185.
- Lekht I, Gulati M, Nayyar M, Katz MD, Ter-Oganesyan R, Marx M, Cen SY, Grant E. Role of contrast-enhanced ultrasound (CEUS) in evaluation of thermal ablation zone. *Abdom Radiol* 2016;41:1511–1521.
- Liu H, Jiang Y-X, Liu J-B, Zhu Q-L, Sun Q. Evaluation of breast lesions with contrast-enhanced ultrasound using the microvascular imaging technique: Initial observations. *Breast* 2008;17:532–539.

- Ma X, Liu R, Zhu C, Zhang J, Ling W. Diagnostic Value of Contrast-Enhanced Sonography for Differentiation of Breast Lesions. *J Ultrasound Med* 2016;35:2095–2102.
- Ma X, Zhang B, Ling W, Liu R, Jia H, Zhu F, Wang M, Liu H, Huang J, Liu L. Contrast-enhanced sonography for the identification of benign and malignant thyroid nodules: Systematic review and meta-analysis. *J Clin Ultrasound* 2015;44:199–209.
- Miyamoto Y, Ito T, Takada E, Omoto K, Hirai T, Moriyasu F. Efficacy of Sonazoid (Perflubutane) for Contrast-Enhanced Ultrasound in the Differentiation of Focal Breast Lesions: Phase 3 Multicenter Clinical Trial. *Am J Roentgenol* 2014;202:W400–W407.
- Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Medicine* 2009;151:264–269.
- Ricci P, Cantisani V, Ballesio L, Pagliara E, Sallusti E, Drudi F, Trippa F, Calascibetta F, Erturk S, Modesti M, Passariello R. Benign and Malignant Breast Lesions: Efficacy of Real Time Contrast-Enhanced Ultrasound vs. Magnetic Resonance Imaging. *Ultraschall Med* 2007;28:57–62.
- Sadigh G, Carlos RC, Neal CH, Wojcinski S, Dwamena BA. Impact of breast mass

- size on accuracy of ultrasound elastography vs. conventional B-mode ultrasound: a meta-analysis of individual participants. *Eur Radiol* 2012;23:1006–1014.
- Schwab F, Redling K, Siebert M, Schötzau A, Schoenenberger C-A, Zanetti-Dällenbach R. Inter- and Intra-Observer Agreement in Ultrasound BI-RADS Classification and Real-Time Elastography Tsukuba Score Assessment of Breast Lesions. *Ultrasound Med Biol* 2016;42:2622–2629.
- Sehgal CM, Arger PH, Rowling SE, Conant EF, Reynolds C, Patton JA. Quantitative vascularity of breast masses by Doppler imaging: regional variations and diagnostic implications. *J Ultrasound Med* 2000;19:427–440.
- Sorelli PG, Cosgrove DO, Svensson WE, Zaman N, Satchithananda K, Barrett NK, Lim AKP. Can contrast-enhanced sonography distinguish benign from malignant breast masses? *J Clin Ultrasound* 2010;38:177–181.
- Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;65:87–108.
- Wang X, Xu P, Wang Y, Grant EG. Contrast-enhanced ultrasonographic findings of different histopathologic types of breast cancer. *Acta Radiologica* 2011;52:248–255.
- Whiting PF. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med* 2011;155:529–536.

Xiao X, Dong L, Jiang Q, Guan X, Wu H, Luo B. Incorporating Contrast-Enhanced Ultrasound into the BI-RADS Scoring System Improves Accuracy in Breast Tumor Diagnosis: A Preliminary Study in China. *Ultrasound Med Biol* 2016;42:2630–2638.

Xiao X, Ou B, Yang H, Wu H, Luo B. Breast Contrast-Enhanced Ultrasound: Is a Scoring System Feasible? ----A Preliminary Study in China. *PLoS ONE* 2014;9:e105517–11.

Xu H-X, Lu M-D. The current status of contrast-enhanced ultrasound in China. *J Med Ultrasonics* 2010;37:97–106.

Zhang J-X, Cai L-S, Chen L, Dai J-L, Song G-H. CEUS Helps to Rerate Small Breast Tumors of BI-RADS Category 3 and Category 4. *Biomed Res* 2014;2014:1–8.

Zhao H, Xu R, Ouyang Q, Chen L, Dong B, Huihua Y. Contrast-enhanced ultrasound is helpful in the differentiation of malignant and benign breast lesions. *Eur J Radiol* 2010;73:288–293.

Zhi H, Ou B, Luo B-M, Feng X, Wen Y-L, Yang H-Y. Comparison of Ultrasound Elastography, Mammography, and Sonography in the Diagnosis of Solid Breast Lesions. *J Ultrasound Med* 2007;26:807–815.

Figures

Fig. 1. Selection process for studies included in the meta-analysis

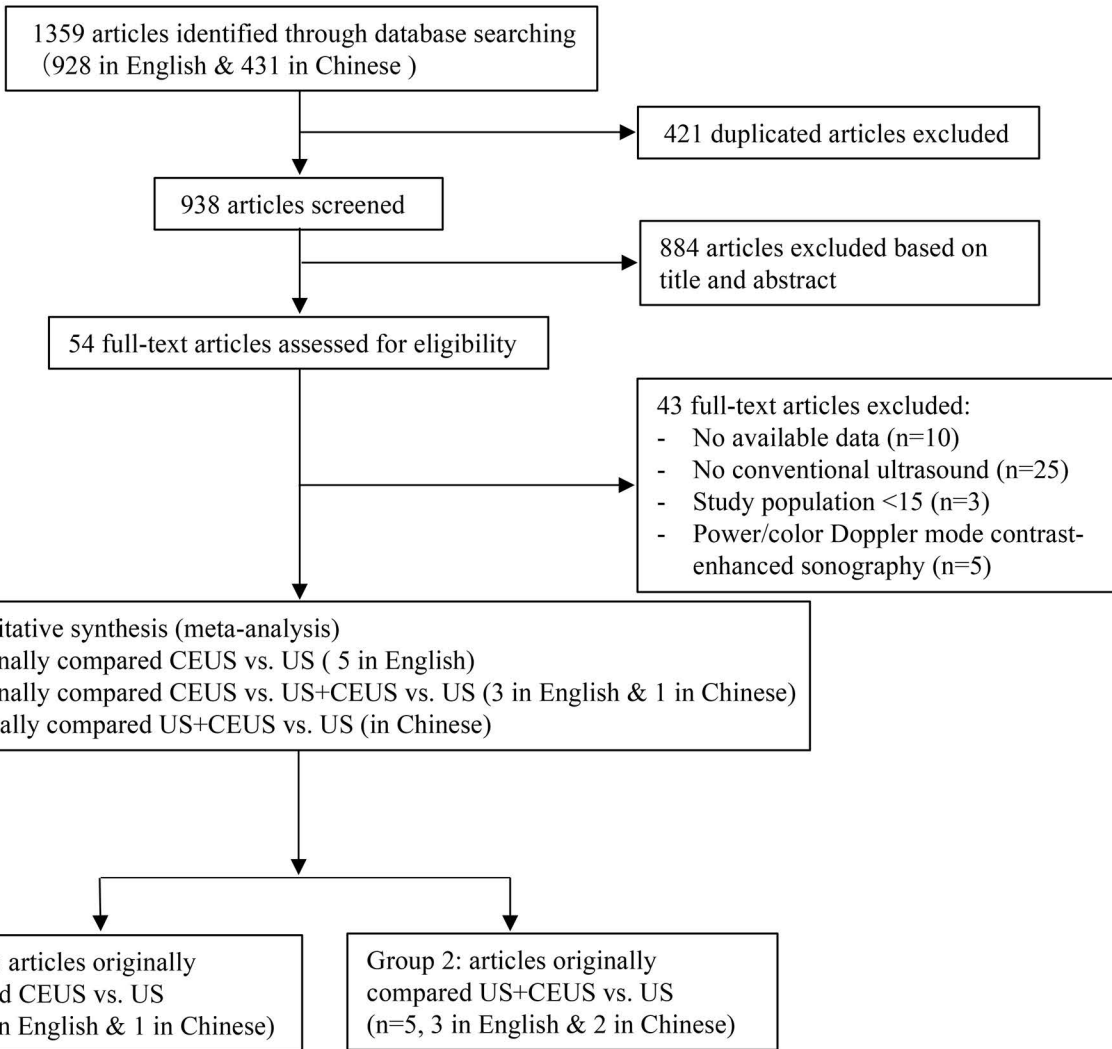
Fig. 2. Risk of bias and applicability concerns graph. Summary of the methodological quality assessment of bias risk and applicability concerns presented for each domain, in percentages across all included studies in the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) -2 criteria. US= ultrasound, CEUS=contrast-enhanced ultrasound.

Fig. 3. Forest plot of sensitivity and specificity of CEUS and US to differentiate benign from malignant breast lesions in studies of Group 1 (CEUS vs. US studies). CEUS=contrast-enhanced ultrasound, US= ultrasound.

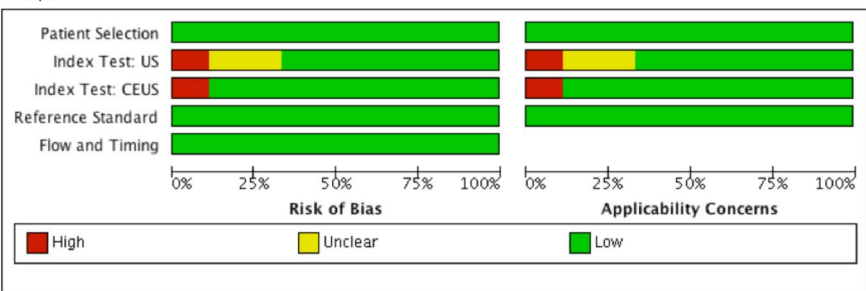
Fig. 4. SROC of CEUS and US to differentiate benign from malignant breast lesions in studies in Group 1 (CEUS vs. US). CEUS=contrast-enhanced ultrasound, US= ultrasound, SROC= summary receiver operating characteristic, AUC=area under the curve, SE=standard error, $Q^*=Q$ index.

Fig. 5. Forest plot of sensitivity and specificity of CEUS and US to differentiate benign from malignant breast lesions in studies in Group 2 (US+CEUS vs. US studies). US+CEUS=ultrasound combined with contrast-enhanced ultrasound, US=ultrasound.

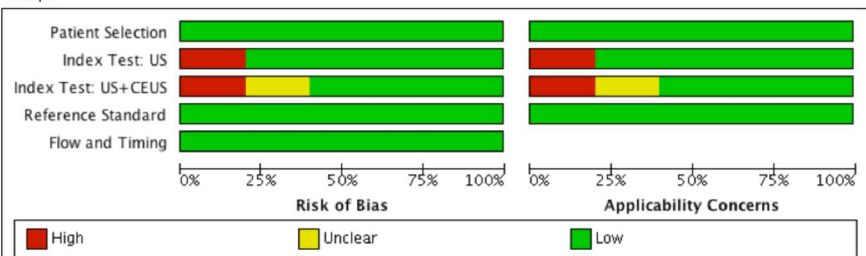
Fig. 6. SROC of US+CEUS and US to differentiate benign from malignant breast lesions in studies of Group 2 (US+CEUS vs. US studies). US+CEUS=ultrasound combined with contrast-enhanced ultrasound, US=ultrasound, SROC= summary receiver operating characteristic, AUC=area under the curve, SE=standard error, $Q^*=Q$ index.



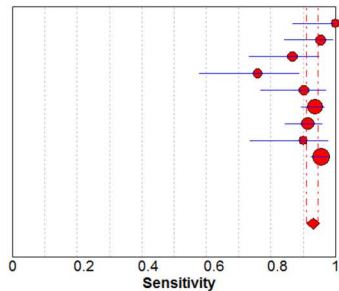
Group 1



Group 2



CEUS

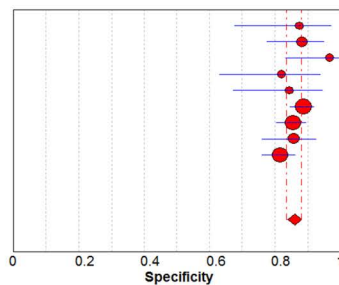


Pooled Sensitivity = 0.93 (0.91 to 0.95)
 Chi-square = 20.64; df = 8 (p = 0.0082)
 Inconsistency (I-square) = 61.2 %

Sensitivity (95% CI)

Ricci et al.	1.00	(0.87 - 1.00)
Liu et al.	0.95	(0.84 - 0.99)
Zhao et al.	0.87	(0.73 - 0.95)
Du et al.	0.76	(0.58 - 0.89)
An et al.	0.90	(0.77 - 0.97)
Xiao et al 1.	0.94	(0.90 - 0.97)
Miyamoto et al.	0.91	(0.84 - 0.96)
Zhang et al.	0.90	(0.73 - 0.98)
xiao et al 2.	0.96	(0.92 - 0.98)

CEUS

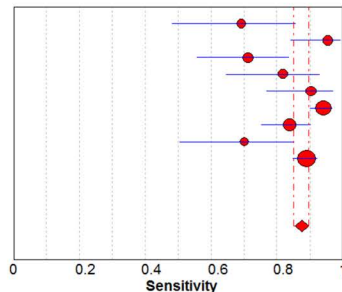


Pooled Specificity = 0.86 (0.84 to 0.88)
 Chi-square = 10.31; df = 8 (p = 0.2443)
 Inconsistency (I-square) = 22.4 %

Specificity (95% CI)

Ricci et al.	0.88	(0.68 - 0.97)
Liu et al.	0.88	(0.77 - 0.95)
Zhao et al.	0.97	(0.83 - 1.00)
Du et al.	0.82	(0.63 - 0.94)
An et al.	0.84	(0.67 - 0.95)
Xiao et al 1.	0.89	(0.84 - 0.92)
Miyamoto et al.	0.85	(0.80 - 0.90)
Zhang et al.	0.86	(0.76 - 0.93)
xiao et al 2.	0.82	(0.76 - 0.86)

US

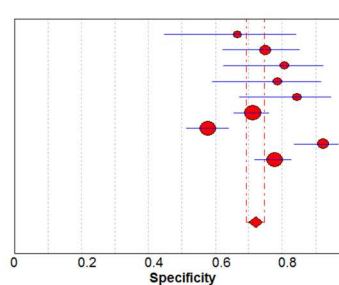


Pooled Sensitivity = 0.87 (0.85 to 0.90)
 Chi-square = 37.32; df = 8 (p = 0.0000)
 Inconsistency (I-square) = 78.6 %

Sensitivity (95% CI)

Ricci et al.	0.69	(0.48 - 0.86)
Liu et al.	0.95	(0.84 - 0.99)
Zhao et al.	0.71	(0.56 - 0.84)
Du et al.	0.82	(0.65 - 0.93)
An et al.	0.90	(0.77 - 0.97)
Xiao et al 1.	0.94	(0.90 - 0.97)
Miyamoto et al.	0.84	(0.75 - 0.90)
Zhang et al.	0.70	(0.51 - 0.85)
Xiao et al 2.	0.89	(0.85 - 0.92)

US

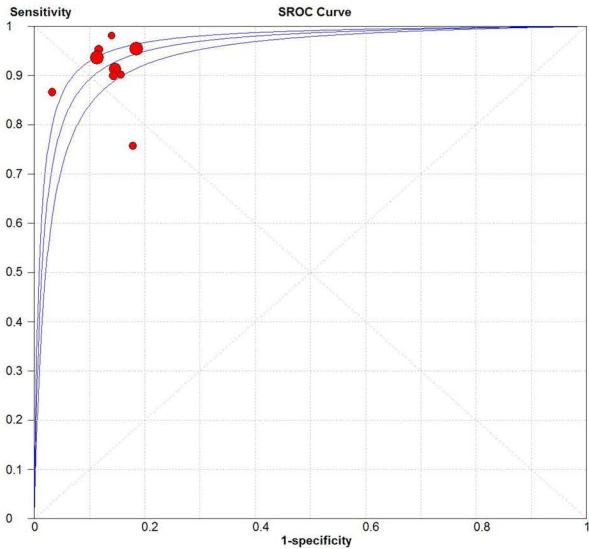


Pooled Specificity = 0.72 (0.69 to 0.75)
 Chi-square = 51.94; df = 8 (p = 0.0000)
 Inconsistency (I-square) = 84.6 %

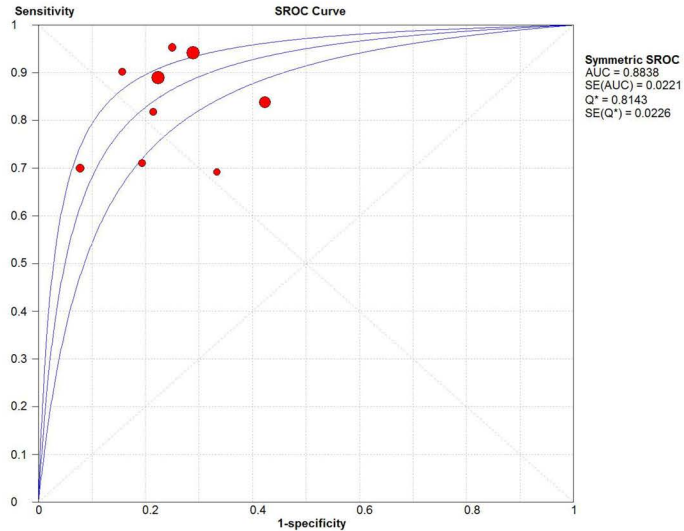
Specificity (95% CI)

Ricci et al.	0.67	(0.45 - 0.84)
Liu et al.	0.75	(0.62 - 0.85)
Zhao et al.	0.81	(0.63 - 0.93)
Du et al.	0.79	(0.59 - 0.92)
An et al.	0.84	(0.67 - 0.95)
Xiao et al 1.	0.71	(0.66 - 0.76)
Miyamoto et al.	0.58	(0.51 - 0.64)
Zhang et al.	0.92	(0.84 - 0.97)
Xiao et al 2.	0.78	(0.72 - 0.83)

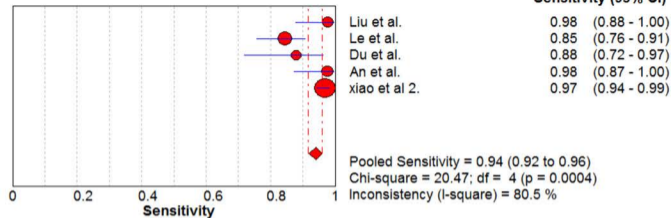
CEUS



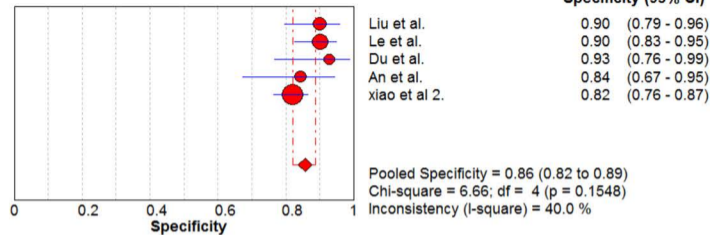
US



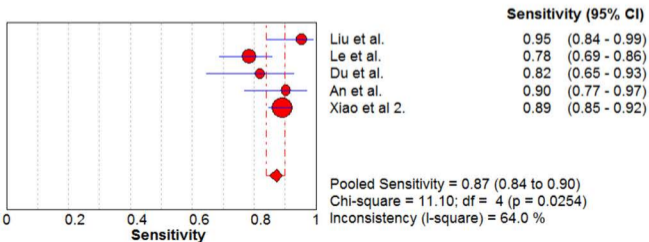
US+CEUS



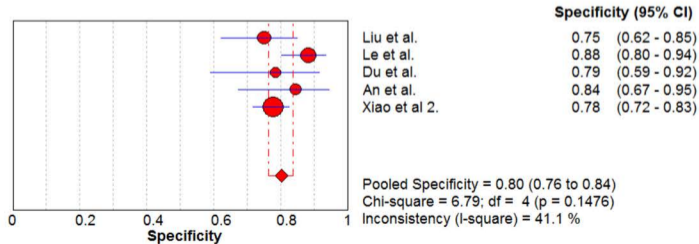
US+CEUS



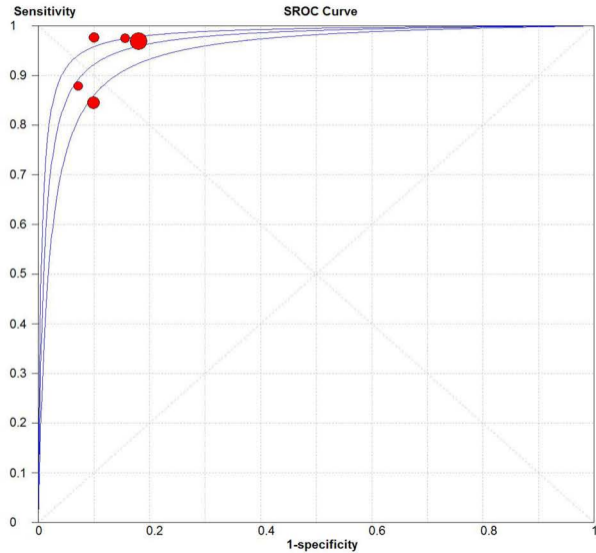
US



US



CEUS + US



US

