The role of grey-scale ultrasound in the diagnosis of adhesive capsulitis of the shoulder: a systematic review and meta-analysis

Huaiyu Wu*, Hongtian Tian*, Fajin Dong, Weiyu Liang, Di Song, Jieying Zeng, Zhimin Ding, Yang Shi, Hui Luo, Jinfeng Xu

* the authors shared the first authorship

Department of Ultrasound, First Affiliated Hospital of Southern University of Science and Technology, Second Clinical College of Jinan University, Shenzhen Medical Ultrasound Engineering Center, Shenzhen People’s Hospital, Shenzhen, Guangdong China

Abstract

Aims: In this systematic review and meta-analysis, we discuss the value of grey-scale ultrasonography (US) in diagnosing adhesive capsulitis of the shoulder (ACS). Material and methods: We retrieved relevant studies from PubMed, Cochrane Library, and Embase before 8 April 2019. We selected 7 studies concerning 446 patients (490 shoulders) that used grey-scale US to diagnose ACS and magnetic resonance imaging (MRI) or arthroscopy as the reference standard. We assessed the diagnostic accuracy of US on the basis of combined sensitivity, specificity, likelihood ratio (LR), and the area under the summary receiver operating characteristic (SROC) curve (AUC). Results: The combined sensitivity, specificity, positive LR and negative LR were found to be 88% (95%CI: 74–95), 96% (95%CI: 88–99), 23.89 (95%CI: 6.31–90.51) and 0.12 (95%CI: 0.05–0.29), respectively. The AUC was 0.97 (95%CI: 0.96–0.98). ACS was diagnosed on the basis of four US features: coracohumeral ligament thickening, inferior capsule/axillary recess capsule thickening, rotator interval abnormality, and restriction of the range of motion. The corresponding sensitivities were 64.4 (95%CI: 48.8–78.1), 82.1 (95%CI: 73.8–88.7), 82.6 (95%CI: 74.1–89.2) and 94.3 (95%CI: 84.3–98.8), respectively, and specificities were 88.9 (95%CI: 76.0–96.3), 95.7 (95%CI: 90.3–98.6), 93.9 (95%CI: 89.8–96.7), and 90.9 (95%CI: 75.7–98.1), respectively. Conclusions: Our meta-analysis showed that grey-scale US plays a significant role in the diagnosis of ACS. Because of its high sensitivity and specificity, US can be added to the existing clinical diagnosis program.

Keywords: ultrasound; adhesive capsulitis; shoulder; systematic review; meta-analysis

Introduction

Frozen shoulder, also known as adhesive capsulitis of the shoulder (ACS), is currently considered as a progressive pain syndrome of the shoulder that is accompanied by capsular fibrosis; it leads to the restriction of the active and passive range of motion (ROM) of the shoulder joint [1-3]. The risk of ACS is high among patients aged 50–70 years, and it escalates with the increase in age [4-5]. ACS seriously affects the patient’s quality of life and it is easily confused with tendinitis, rotator cuff tear, arthritis, etc, which require a totally different treatment [6]. Good prognosis can be achieved by correct physical therapy or invasive treatment followed by rehabilitation training [7]. Therefore, early and accurate diagnosis of ACS is particularly important. At present, the imaging techniques used for diagnosing ACS include plain radiography, arthrography, ultrasound (US) and magnetic resonance imaging (MRI)/magnetic resonance arthrography (MRA) [6,8-11].

Currently, musculoskeletal diagnostic ultrasound imaging (MSK-DUSI) technique is advancing rapidly and has become an important imaging method for diagnos-
The role of grey-scale ultrasound in the diagnosis of adhesive capsulitis of the shoulder

The most important advantage of high-resolution US is that anatomical structures such as the coracohumeral ligament (CHL), inferior capsule/axillary recess capsule (IC/ARC) and rotator interval (RI) can be clearly and dynamically observed. In addition, US is an inexpensive and non-invasive technique, and it can be widely used in patients with non-MRI-compatible implants. US can not only help diagnose ACS, but also guide intra-articular corticosteroid injections, which can effectively control inflammation [13-15].

Plain radiography has little value in the diagnosis of ACS. It is mainly used to identify calcific tendinopathy [4]. Currently, MRI/MRA and arthroscopy are widely used to diagnose ACS. These techniques show high sensitivity and specificity [16-18] but, the techniques have low accessibility, are invasive and expensive. There are few studies on the application of US for diagnosing ACS. Therefore, a meta-analysis is required to evaluate the efficiency of this technique and to help clinicians make a correct diagnosis using quantitative data.

In this study, we aimed to determine the accuracy of US, relative to that of MRI/MRA or arthrography, in diagnosing ACS using combined sensitivity, specificity, likelihood ratio (LR) and summary receiver operating characteristic (SROC) and to determine whether US can be added to the existing clinical diagnostic plan.

Material and methods

We incorporated the studies that used grey-scale US to diagnose ACS and provided, directly or indirectly, raw data for calculating sensitivity, specificity, positive LR (LR+), negative LR (LR−), and area under SROC curve (AUC). We limited our search to literature published in English and excluded studies with no reference standards, studies with a small sample size (<20 cases), incomplete studies, reviews, case reports, systematic reviews and meta-analyses.

We identified and retrieved all relevant English literature from PubMed, Embase, and the Cochrane library. The search terms used include Medical Subject Headings (MeSH) and free words (details in supplementary file on the journal site).

Two authors (HYW and HTT) independently read the titles and abstracts of the retrieved studies and excluded duplicates and studies that did not meet the inclusion criteria. Subsequently, they read the remaining articles completely to determine their true eligibility and cross-checked the results of the included studies. If they were controversial, they discussed with each other or with a third party (JFX). Some articles did not provide sufficient information; therefore, the authors were contacted to obtain relevant information, which helped decide the articles’ eligibility.

After the inclusion of eligible articles, the two authors (HYW and FJD) determined the following standards for extracting information: the first author’s surname, year of publication, study design, number of patients, reference standard, age of ACS and non-ACS patients, true positivity (TP), false negativity (FN), false positivity (FP), true negativity (TN), sensitivity, and specificity.

The quality of the included literature was evaluated using the QUADAS-2 tool (quality assessment of diagnostic accuracy studies-2) [19]. The tool is composed of 4 bias-related domains and 3 applicability concerns. The 4 domains were patient selection, index test, reference standard, and flow and timing. The applicability concerns were structured similar to bias-related domains, but they did not include flow and timing. All domains are evaluated for bias and risk and classified as low, high, or unclear. If all the bias-related domains or applicability concerns were judged as “low,” the study was considered to have a “low risk of bias” or “low concern regarding applicability.” If one or more domains or concerns were judged as “high” or “unclear,” the study may be considered to have a “risk of bias” or “concerns regarding applicability.”

From the included studies, the data on TP, TN, FP, and FN were extracted using a fourfold table. Data were statistically analysed using RevMan 5.3, Stata 12.0, and meta-disc 1.4.0. RevMan 5.3 was used to assess the methodological quality of the eligible studies. Stata 12.0 was used to pool the statistical indices from all the studies and to draw statistical graphs, and meta-disc 1.4.0 was used to pool the different statistical indices of US features, such as the pooled forest graph of sensitivity, specificity, LR+, and LR− with 95% confidence intervals (CI) and SROC-AUC.

Heterogeneity of the included literature was evaluated. The type of statistical model was selected according to the results of the heterogeneity test. Heterogeneity analysis of the results of each study was performed using the chi-square test, and heterogeneity was evaluated on the basis of I². For example, if I² is <50%, which indicates that there is no statistical heterogeneity between the studies, the fixed effect model is used to estimate the combined effect. If I² is ≥50%, which indicates that there is statistical heterogeneity among the studies, it is necessary to analyse the causes of heterogeneity. If the heterogeneity cannot be excluded, the random effect model can be considered to estimate the combined effect [20].

Deek’s funnel plot asymmetry test was used to evaluate potential publication bias. It was conducted by the regression of diagnostic log odds ratio (lnDOR) versus the
inverse of the square root (1/sqrt) of the effective sample size (1/ESS\(^{1/2}\)) and weighted by the effective sample size. A p-value of <0.10 for the slope coefficient indicated significant asymmetry [21].

**Results**

Figure 1 outlines the process used for selecting the studies. According to the search strategy, a total of 5,021 articles - 1,930, 3,126, and 145 articles from PubMed, Embase, and the Cochrane library, respectively - were extracted as of 8 August 2019. Duplicate articles (284) were eliminated using ENDNOTE X7.4. Among the remaining 4917 articles, 4,799 articles including 31 systematic reviews or meta-analyses (or both) and 418 case reports or reviews (or both) were excluded because the title and abstract were not relevant to the current meta-analysis. Among the remaining 118 articles, 111 were excluded because they did not provide valid data. Finally, 7 articles were included. They included a total of 446 patients (490 shoulders). Table I and Table II show details of the finally included articles.

Figure 2 shows the quality of the included literature. In general, the results of this show whether the studies are reliable.

Firstly, six articles [22-25,27,28] were associated with a high risk of bias with regard to patient selection, because they were all controlled trials and did not mention whether patients were consecutively enrolled or not. The other research [26] did not mention whether the included patients were consecutive too, but a case-control design was avoided, so it was associated with an unclear risk.

Secondly, four articles [22,26-28] were associated with a low risk with regard to domain 2, index test. Sonographers already knows the diagnosis of patients in 4 studies [23-25]; therefore, these studies are associated with a high risk of bias.

Thirdly, five articles [22,23,25,27,28] were associated with a low risk with regard to domain 3. Authors of two articles [24,26] pointed out that, in their studies, the

Table I. Characteristics of the included studies

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Country</th>
<th>Design</th>
<th>No. of patients (No. of shoulders)</th>
<th>No. of ACS (Non-ACS) cases</th>
<th>Criteria</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheng, 2017 [22]</td>
<td>China</td>
<td>Retrospective, control</td>
<td>90 (90)</td>
<td>45 (45)</td>
<td>CHL, IC, RI</td>
<td>Clinical, MRI</td>
</tr>
<tr>
<td>Kim, 2018 [23]</td>
<td>South Korea</td>
<td>Retrospective, control</td>
<td>44 (88)</td>
<td>44 (44)</td>
<td>ARC (IC)</td>
<td>Clinical</td>
</tr>
<tr>
<td>Lee, 2005 [24]</td>
<td>UK</td>
<td>Prospective, control</td>
<td>140 (140)</td>
<td>30 (110)</td>
<td>RI</td>
<td>Clinical, Arthroscopic</td>
</tr>
<tr>
<td>Ryu, 1993 [26]</td>
<td>South Korea</td>
<td>Retrospective, cohort</td>
<td>26 (26)</td>
<td>23 (3)</td>
<td>ROM</td>
<td>Clinical, Arthrographic</td>
</tr>
<tr>
<td>Sernik, 2019 [27]</td>
<td>Brazil</td>
<td>Prospective, control</td>
<td>51 (51)</td>
<td>23 (28)</td>
<td>ARC (IC)</td>
<td>MRI</td>
</tr>
<tr>
<td>Tandon, 2017 [28]</td>
<td>India</td>
<td>Prospective, control</td>
<td>60 (60)</td>
<td>30 (30)</td>
<td>CHL, RI, ROM</td>
<td>Clinical, MRI</td>
</tr>
</tbody>
</table>

No., number; ACS, Adhesive Capsulitis of the Shoulder; CHL, coracohumeral ligament; IC, inferior capsule; RI, Rotator interval; ARC, axillary recess capsule; ROM, restriction of the range of motion

**Fig 1.** Flow chart outlining the process of study selection.
The role of grey-scale ultrasound in the diagnosis of adhesive capsulitis of the shoulder

reference standard results were interpreted on the basis of the results of the index tests; therefore, these studies are associated with a high risk of bias.

Lastly, one article [28] was associated with a high risk with regard to domain 4; while 4 articles [23-25,27] were associated with unclear risk and 2 [22,26] were associated with a low risk. Tandon et al [28] did not include all the enrolled patients in the analysis. Four studies [23-25,27] did not mention the interval between the index test and reference standard. In addition, the applicability concerns with regard to domains 1, 2, and 3 are low.

Among all the results, sensitivity, specificity, LR+ and LR− with 95% CI and forest plots (fig 3a–d) showed obvious heterogeneity, with an I² of 81.20%, 87.72%, 75.08%, and 84.36%, respectively. Therefore, we used the random effects model to merge the included documents. The combined results are shown in Figure 3a–e. The combined sensitivity was 0.88 (95% CI: 0.74–0.95) and specificity was 0.96 (95% CI: 0.88–0.99). LR+ and LR− were 23.89 (95% CI: 6.31–90.51) and 0.12 (95% CI: 0.05–0.29), respectively and the post-test probability is 86% and 3%, respectively. Figure 3f shows ROC curves for the US-based diagnosis of ACS. The AUC is 0.97 (95% CI:0.96–0.98). Figure 3g shows Deek’s funnel plot (p=0.504). The included studies were basically distributed on both sides of the regression line; therefore, the publication bias was not obvious.

Table II. Characteristics of the included studies

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>CHL thickening</th>
<th>IC/ARC thickening</th>
<th>RI abnormality</th>
<th>ROM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cut-off (mm)</td>
<td>Sen (%)</td>
<td>Spe (%)</td>
<td>Cut-off (mm)</td>
</tr>
<tr>
<td>Cheng, 2017 [22]</td>
<td>3.0</td>
<td>64.4</td>
<td>88.9</td>
<td>3.5</td>
</tr>
<tr>
<td>Kim, 2018 [23]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.2</td>
</tr>
<tr>
<td>Lee, 2005 [24]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Lee, 2010 [25]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ryu, 1993 [26]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.0</td>
</tr>
<tr>
<td>Sernik, 2019 [27]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7</td>
</tr>
<tr>
<td>Tandon, 2017</td>
<td>0.7</td>
<td>64.4</td>
<td>88.9</td>
<td>88.9</td>
</tr>
<tr>
<td></td>
<td>15.22 (6.70–34.59)</td>
<td>0.15 (0.04–0.58)</td>
<td>15.22 (6.70–34.59)</td>
<td>0.15 (0.04–0.58)</td>
</tr>
</tbody>
</table>

ACS, Adhesive Capsulitis of the Shoulder; CHL, coracohumeral ligament; IC, inferior capsule; ARC, axillary recess capsule; ROM, restriction of the range of motion; Sen, Sensitivity; Spe, Specificity; LR+, positive LR; LR−, negative LR; AUC, the area under the summary receiver operating characteristic curve

Table III. Characteristics of the four US features.

<table>
<thead>
<tr>
<th>US feature</th>
<th>No. of studies</th>
<th>Sen (%)</th>
<th>Spe (%)</th>
<th>LR+</th>
<th>LR−</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHL</td>
<td>1</td>
<td>64.4 (48.8–78.1)</td>
<td>88.9 (76.0–96.3)</td>
<td>5.8 (-)</td>
<td>0.4 (-)</td>
<td>-</td>
</tr>
<tr>
<td>IC/ARC</td>
<td>3</td>
<td>82.1 (73.8–88.7)</td>
<td>95.7 (90.3–98.6)</td>
<td>15.22 (6.70–34.59)</td>
<td>0.15 (0.04–0.58)</td>
<td>98.0</td>
</tr>
<tr>
<td>RI</td>
<td>4</td>
<td>82.6 (74.1–89.2)</td>
<td>93.9 (89.8–96.7)</td>
<td>12.59 (2.33–68.15)</td>
<td>0.17 (0.07–0.43)</td>
<td>86.9</td>
</tr>
<tr>
<td>ROM</td>
<td>2</td>
<td>94.3 (84.3–98.8)</td>
<td>90.9 (75.7–98.1)</td>
<td>9.25 (3.43–24.99)</td>
<td>0.08 (0.02–0.35)</td>
<td>-</td>
</tr>
</tbody>
</table>

US, Ultrasound; No., number; CHL, coracohumeral ligament; IC, inferior capsule; ARC, axillary recess capsule; RI, Rotator interval; ROM, restriction of the range of motion; Sen, Sensitivity; Spe, Specificity; LR+, positive LR; LR−, negative LR; AUC, the area under the summary receiver operating characteristic curve
In the articles included in this study, the four US features, namely CHL, IC, RI and ROM, were used/evaluated at varying degrees. The combined sensitivities and specificities of these four US features are presented in Table III, along with combined LR+, LR− and AUC.

Discussion

At present, because the traditional diagnostic methods are inefficient in diagnosing ACS accurately, patients have been losing the opportunity of timely physical therapy and an improved quality of life [29,30]. As of 8 April 2019, the databases PubMed, Cochrane library and Embase have not shown any systematic reviews and meta-analyses on the application of US as a diagnostic method for ACS.

Therefore, the purpose of this meta-analysis was to evaluate the diagnostic value of US for ACS. After analysis, the combination sensitivity, specificity, LR, and SROC-AUC were obtained. Although these indicators showed high heterogeneity ($I^2 \geq 50\%$) that cannot be ignored, US still shows high sensitivity, specificity, LR+, and AUC and lower LR−. LR+ and LR−, which describe the property of discrimination between positive and negative test results, respectively, were calculated [31]. The AUC of ROC reflects the accuracy of the diagnostic test. AUC values indicate the following: 0.5–0.7, low accuracy; 0.7–0.9, moderate accuracy; and >0.9, high accuracy [32]. Therefore, US shows high accuracy in the diagnosis of ACS.

At present, ACS is diagnosed using grey-scale US on the basis of the CHL, IC/ARC, RI, and ROM. The CHL, IC/ARC, and RI are thicker and the ROM is lower in patients with ACS than in those without ACS [33-35]. This study separately analysed the diagnostic value of these 4 features. The IC/ARC, RI, and ROM were found to have a very high diagnostic value, which was no less than that of MRI/MRA [16]. The sensitivity of ROM was found to be the highest in this study (94.3, 95% CI: 84.3–98.8); restriction of external rotation of the shoulder showed a statistically significant diagnostic value (p=0.0001) [28]. The IC/ARC of patients with ACS shows the highest

Fig 3. A-G Sensitivity (A), specificity (B), LR+ (C) and LR−(D) with 95% CI and forest plots, and the post-test probability (E), SROC(F) and Deek’s funnel (G) of grey scale ultrasound diagnosis of adhesive capsulitis of the shoulder.
specificity (95.7, 95% CI: 90.3-98.6), probably because its thickening is an important anatomical and pathophysiological abnormality of ACS [36]. Significantly, in the study, although the sensitivity of ROM (94.3%) is higher than the combined sensitivity (88.0%) in using gray-scale US to diagnose ACS, this did not necessarily affect the diagnostic value of US in ACS. The lower combined sensitivity may be due to the fact that ROM was not used to diagnose ACS in gray-scale US for most previous studies. Among the included studies, only one study used CHL thickness to diagnose ACS. Another included study [28] discussed the use of CHL thickness, but it did not provide fourfold data; therefore, it could not be analysed. In fact, there are many studies that used changes in the CHL in the diagnosis of ACS [33,37]; however, none of them computed the accuracy, sensitivity, and specificity. Furthermore, these studies have shown that US is valuable in diagnosing ACS and proved that CHL thickening is valuable in diagnosing ACS. Recently, studies that applied grey-scale US technology for the diagnosis of ACS have emerged internationally [22,37-39]; these studies have greatly improved doctors' confidence in the diagnostic accuracy of grey-scale US.

The systematic review process used in this study was rigorous; however, there are limitations to this study. Quality assessment of the 7 documents included indicated that they all present high-risk results. Six of the studies established a control group and there may have been selection bias or partial verification bias. In addition, this systematic review did not collect grey literature; its language restriction was English, and the search strategy was not detailed enough. At present, there are few documents on the application of US in the diagnosis of ACS and most of the data are incomplete. Therefore, fewer articles were included and there is a certain level of publication bias, language bias, and search bias. The final results of the study showed that the heterogeneities of combined sensitivity, specificity, LR+, and combined LR− were all high, which may be related to the study design and the US features used in the study.

Our results may have important clinical implications because US is more widely available, convenient to use and cost efficient for shoulder joint examinations. It can facilitate the timely detection of other lesions accompanying ACS [40-42] and the determination of a treatment plan as early as possible. Because many shoulder joint disorders limit the range of motion, the use of clinical criteria to diagnose ACS alone may lead to the neglect of other shoulder joint-related diseases, such as rotator cuff tears. In addition, immediate treatment (such as physiotherapy or intra-articular steroid injection) of ACS after diagnosis can help shorten the duration of joint stiffness and achieve a good prognosis.

However, there are certain limitations to using US for diagnosing ACS. In China, orthopaedic surgeons are more dependent on MRI and the popularity of MSK-DUSI is far from enough. Only experienced and highly trained sonographers can perform shoulder-dynamic and static US scans. Studies have shown that once the dynamic and static US are combined, the sensitivity of diagnosing ACS can be as high as 100%, but the specificity may be reduced to 86.7% [28]. Complex inspection techniques are difficult to perform in countries with a large population such as China.

Therefore, our team believes that when a patient is clinically suspected to be more likely to have ACS, detailed US may be preferred if the hospital has the MSK-DUSI facility. ACS is highly suspected once the US is positive because of its very high sensitivity. It should be emphasized that once US is negative, but the patients is still clinically suspected to have ACS, the possibility of ACS cannot be easily ruled out; other imaging methods or US technologies may be needed to validate the results. Therefore, although US-based diagnosis of ACS shows high sensitivity and specificity at the same time, using US-based exclusion of ACS in clinical practice remains to be further verified.

Conclusions

Our study shows that US is very important for diagnosing ACS. Because it is characterized by high sensitivity and specificity, convenience, non-invasive nature, and guided puncture, US-based examination of ACS can be added to existing clinical diagnostic protocols.

Acknowledgment

We thank Zongyu Zhang and Jinsen Zou at the Shenzhen People's Hospital for their encouragement. This work was supported by the National Natural Science Foundation of China (Grant No. 81771841) and the Commission of Scientific and Technology of Shenzhen (Grant No. JCYJ20170307095706970).

References

34. Michelin P, Delarue Y, Duparc F, Dacher JN. Thickening of the supraspinatus tendon during dynamic ultrasonography can predict a decrease in capacity and MR arthrographic