REVIEW

Nr 2021;11 (3):383-408

Diagnostic Accuracy of Physical Tests and Imaging Techniques in Patients with Shoulder Impingement Syndrome

M. Lorusso[1,2], E. Mastrangelo[3], G. Garofalo[4], D. Ristori[5], F. Brindisino[1,6]

[1] *Department of Clinical Sciences and Translational Medicine, Tor Vergata University of Rome, Rome, Italy*
[2] *"Equilibrio" Physiotherapy Clinic, Andria, Barletta-Andria-Trani, Italy*
[3] *Physiotherapist in Bari, Bari, Italy*
[4] *Physiotherapist in Lioni, Avellino, Italy*
[5] *Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genova, Campus of Savona, Genoa, Italy*
[6] *Department of Medicine and Health Science Vincenzo Tiberio, University of Molise, Campobasso, Italy*

CORRESPONDING AUTHOR:
Mariangela Lorusso
"Equilibrio" Physiotherapy Clinic
76123 Andria
Barletta-Andria-Trani, Italy
E-mail: mariangela.lorusso@hotmail.it

DOI:
10.32098/mltj.03.2021.03

LEVEL OF EVIDENCE: 1A

SUMMARY

Background. A common cause of shoulder pain is attributable to the impingement syndrome. The objectives of this systematic review are: 1) to investigate the usefulness of clinical tests and their psychometric qualities in the clinical diagnosis of subacromial shoulder impingement; 2) to investigate the accuracy of different imaging methods to identify structural correlates related to rotator cuff injuries as indirect confirmation of subacromial impingement syndrome.

Materials and methods. This systematic review was carried out following the indications contained in the PRISMA checklist. The methodological quality of the studies included was assessed through the AMSTAR 2.

Results. Several studies agree that no test improves the post-test probability of detecting subacromial impingement and, therefore, are inaccurate. Ultrasound, Magnetic Resonance Imaging and Magnetic Resonance Arthrography have the same level of sensitivity and specificity in the detection of complete lesions of the rotator cuff. For partial injuries it seems that there is a lower sensitivity in view of high specificity, especially with ultrasound.

Discussion and conclusions. The present review confirms the poor diagnostic ability of clinical tests for subacromial shoulder impingement and it highlights the criticism in imaging employment in the complex process of the clinical framing of this syndrome.

KEY WORDS

Diagnostic imaging; physical examination; shoulder pain; screening; subacromial impingement syndrome.

BACKGROUND

Shoulder pain is the third most common musculoskeletal disorder in the general population and can significantly influence the patient's ability to perform work, daily life, and leisure activities (1-3). Shoulder problems have also a significant social and economic impact (1).

The punctual prevalence of shoulder pain varies between 6.9% and 26% in the adult population and increases with age (1); the annual incidence varies according to age: 0.9% for those aged between 31 and 35, 2.5% for 42-46 years, 1.1% for 56-60 years and 1.6 % for those aged 70 to 74 (1).

A common cause of shoulder pain without joint stiffness is attributable to the subacromial impingement syndrome (SIS), a nosological entity diagnosed in about 74% of patients with painful shoulder; SIS is also reported as a contributing factor to the onset of shoulder pain from 48% to 65% (4-6).
SIS was initially described by Neer as a pinching of soft tissues between the humerus and the coracoacromial arch during the elevation movement of the arm (7); later it was defined as a clinical entity rather than a diagnostic one since it can be associated with many tissue, and functional alterations (8).
Shoulder impingement can be caused by external (external impingement) or internal causes (internal impingement) (9, 10); among the "external impingement", the sub-coracoid and the SIS are included (11). SIS can occur due to different mechanisms and structural factors and functional factors can contribute to its onset (**table I**) (11).
According to Neer (7), SIS is due to primary abnormalities of the coracoacromial arch, such as the acromial spur or hook morphology of the acromion which cause compression of the subacromial bursa and abrasion of the bursal part of the tendon of the supraspinatus (7).
Subsequently, Lewis (12) highlighted the hypothesis that the subacromial conflict, although present, represents the consequence of an underlying incompetence of the rotator cuff which leads to an alteration of the normal biomechanics of stability of the humerus inside the glenoid (12).
The conflict of such structures, then, would only be the tip of the iceberg, one of the last phases of the functional involution of a shoulder with an incompetent rotator cuff, before it is injured (12).
However, if impingement is not impingement, it means the pathophysiology is not the rotator cuff hitting the acromion, but other mechanisms and not acromial contact cause pain in these patients (13). In the light of current knowledge, tendinopathy seems to be the result of a "failed healing response" related to age-related factors, genetic factors, load alterations, trauma, comorbidities, sports, drugs, smoking, unhealthy lifestyle (13).
Through the years, therefore, we passed from an extrinsic conception of pathology, to an intrinsic conception, according to which the rotator cuff tendinopathy originates in the inner layers of the tendon as a consequence of friction between the intratendinous layers, overuse, nonuse or overload (9).
So, the SIS has been defined as a clinical rather than diagnostic entity, which corresponds to a series of tissue and functional alterations (5, 13).
Moreover, SIS in the advanced phase is associated with rotator cuff injuries even though the relationship between these two entities is still a matter of debate (14); however, lesion of the supraspinatus tendon is often identified as an indirect sign of the SIS (4, 13).
At the base of the diagnostic process of the SIS there are the patient history and the physical examination (15). The clinical tests and some instrumental examinations such as Ultrasound (US), Magnetic Resonance Imaging (MRI) and Magnetic Resonance Arthrography (MRA) (see **Appendix 1**) over time have begun to be identified as an integral part of this process, although in Guidelines by Diercks *et al*, imaging were not recommended before 6 weeks after the onset of symptoms (7, 16, 17).
Numerous studies have proposed to investigate the quality of the psychometric properties of the symptom provocation procedures and the validity compared to the use of diagnostic imaging to confirm the diagnosis of SIS in patients with painful shoulder (16-20).
However, methodological quality of the studies on these topics revealed concerns, and there are also few certainties about the use of imaging techniques as a valid aid to discriminate between symptomatic and asymptomatic patients (12, 18, 19).
We will perform two systematic reviews on two different topics in order to allow clinicians to understand the best methods of clinical evaluation and instrumental diagnostics in patients with SIS.
Therefore, the objectives of this systematic review are:

- to investigate the usefulness of clinical tests and their psychometric qualities in the clinical diagnosis of SIS;

Table I. Subacromial impingement syndrome: structural factors and functional factors.

Structural factors	Bursae (inflammation, thickening). Rotator cuff tendon (tendinitis, thickening, partial-thickness tears). Humeral head (congenital abnormalities, fracture malunion). Acromioclavicular joint (joint abnormalities, sprains, degenerative spurs). Acromion (abnormal shape, spurs, os acromiale unfused, malunion of fracture, nonunion of fracture).
Functional factors	Rotator cuff (weakness, inflammation). Imbalance (poor dynamic stabilization). Capsular (hypomobility, hypermobility). Scapular factors (postural adaptations, position, restriction in motion, neuromuscular control, paralysis, facioscapulohumeral muscular dystrophy).

- to investigate the accuracy of different imaging methods to identify structural correlates related to rotator cuff injuries as indirect confirmation of SIS.

MATERIALS AND METHODS

The study meets the ethical standards of the journal (21). This systematic review was carried out following the methodological indications contained in the PRISMA checklist (22). A bibliographic search was conducted from July 2018 to December 2018 on the PubMed database (US National Library of Medicine National Institutes of Health) using a combination of keywords and MeSH terms (**appendix 2**) designed to identify studies on clinical tests and instrumental examination in the diagnosis of SIS.

Appendix 3 reported the search strings used with the filter to select systematic reviews, and eligibility criteria.

Two reviewers (ML, EM) for the clinical test string and (ML, GG) for the instrumental examinations string have independently analysed titles, abstracts and full-text to identify the articles of interest. For each point of conflict, an auditor super-partes (FB) has been consulted.

The methodological quality of the articles included in the study has been assessed independently by two reviewers for each string through the AMSTAR 2 scale. AMSTAR 2 is the new revised checklist that evaluates systematic reviews (23). With the new AMSTAR 2 version, it is possible to evaluate systematic reviews including NRSI (Non-Randomised Studies of Interventions) and Randomized Controlled Trial (RCTs). The checklist includes 16 items with the possibility to answer "YES", "NO", "PARTIAL YES".

The AMSTAR 2 overall score (**appendix 4**) is calculated by entering the answers on the website (https://amstar.ca/Amstar_Checklist.php) (2).

RESULTS

Selection of studies on psychometric qualities of clinical tests

The research conducted on PubMed resulted in 264 articles. By title and abstract screening 249 studies were excluded. 15 articles were selected for full-text analysis and 3 articles were excluded because they did not meet the inclusion criteria. Finally, 12 articles were included, specifically including 7 meta-analysis and 5 systematic reviews (**table II**).

Selection of studies on diagnostic accuracy of instrumental examinations

The research conducted on PubMed produced 170 results. By screening of title and abstract 151 articles were excluded. After reading the full text, 12 articles were excluded. Finally, 7 systematic reviews have been included (**table III**).

The study selection processes for the reliability of the tests and for the accuracy of the imaging assessment are summarised in the PRISMA flow-charts (**figures 1, 2**).

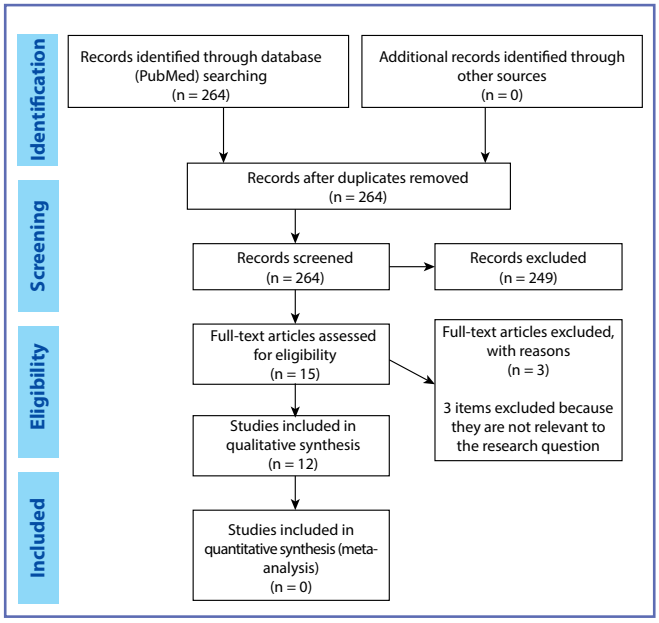

Figure 1. Prisma flow-chart Clinical Tests.

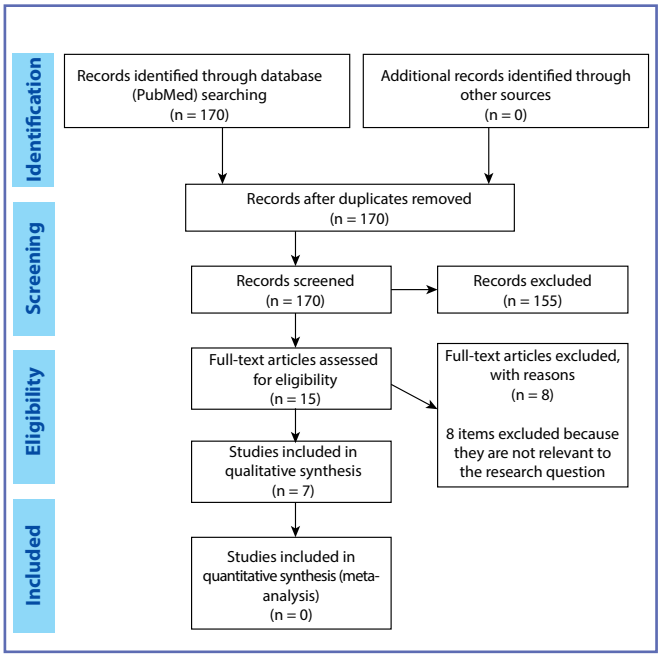

Figure 2. Prisma flow-chart Instrumental Examinations.

Table II. Synoptic table articles included for clinical tests.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Gismervik *et al*., 2017 Physical examination tests of the shoulder: a systematic review and meta-analysis of diagnostic test performance.	Systematic Review and meta-analysis	This meta-analysis aims to use diagnostic odds ratio (DOR) to evaluate how much physical examination tests of the shoulder (PETS) shift overall probability and to rank the test performance of single PETS in order to aid the clinician's choice of which tests to use. .	Intervention: a fixed effect model was used to assess the overall diagnostic validity of PETS by pooling DOR for different PETS with similar biomechanical rationale when possible. Single PETS were assessed and ranked by DOR. Clinical performance was assessed by sensitivity, specificity, accuracy and likelihood ratio.	Hawkins test obtained the highest DOR (2.86) for impingement syndrome (sensitivity 0.58, specificity 0.67). No single PETS showed superior clinical test performance.	The clinical performance of single PETS is limited. The authors suggest that clinicians choose their PETS among those with the highest pooled DOR.
Lange *et al*., 2016 Reliability of specific physical examination tests for the diagnosis of shoulder pathologies: a systematic review and meta-analysis	Systematic Review and meta-analysis	The aim of this systematic review was to summarize and evaluate intra-rater and inter-rater reliability of physical examination tests in the diagnosis of shoulder pathologies	Intervention: the search strategy revealed 3259 articles, of which 18 finally met the inclusion criteria. Pattern: all included studies were prospective and inter-rater reliability was assessed in 17, and 1 study evaluated inter-rater and intra-rater reliability.	The included studies were of low methodological quality according to the QAREL tool. Meta-analysis identified extensive heterogeneity among studies for physical examination tests, thus the findings of the meta-analysis may be inaccurate and need to be interpreted with caution.	This review identified a lack of high- quality studies evaluating inter-rater as well as intra-rater reliability of specific physical examination tests for the diagnosis of shoulder pathologies. In addition, reliability measures differed between included studies hindering proper cross-study comparisons.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
O'Kane *et al.*, 2014 The Evidenced-Based Shoulder Evaluation	Systematic Review and meta-analysis	The purpose of this article is to review the evidence for several physical examination techniques and imaging modalities and to propose an evidenced-based strategy for the evaluation of the painful shoulder.	Intervention: abstracts were screened for statistical assessment of examination or imaging performance. In developing the evaluation algorithm, the few shoulder tests identified as performing well in a recent meta-analysis were included. In addition, a number of studies identified as low bias using the QUADAS-2 tool were reviewed, and their reported sensitivity and specificity were used to calculate likelihood ratios evaluating the role of those tests in the shoulder evaluation algorithm.	In total, 57 named shoulder tests were identified. Meta-analysis was possible for 16 tests assessing 5 diagnoses, as follows: subacromial impingement, superior labral anterior to posterior (SLAP) tear, anterior or posterior labral tear, anterior instability, and rotator cuff tendinopathy.	The literature is replete with studies describing physical examinations for the shoulder, although many are of insufficient quality. Moreover, combining history and physical examination improves accuracy over physical examinations in isolation.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Hermans *et al*., 2013 Does This Patient With Shoulder Pain Have Rotator Cuff Disease? The Rational Clinical Examination Systematic Review	Systematic Review and meta-analysis	To perform a meta-analysis to identify the most accurate clinical examination findings for rotator cuff disease (RCD), including subacromial bursitis and tendinopathy which can lead to a clinical entity known as subacromial impingement syndrome.	Intervention: two reviewers extracted study characteristics and diagnostic accuracy data for the index and reference tests of each study. For each finding, we recalculated the sensitivity, specificity, and likelihood ratios (LRs). Two pairs of reviewers independently assigned levels of evidence.	Twenty-eight studies assessed the examination of referred patients by specialists. Only 5 studies reached Rational Clinical Examination quality scores of level 1-2. Because of the relatively low diagnostic accuracy of commonly performed individual tests, combinations of findings for RCD have been evaluated. However, a positive Hawkins test result together with a positive Neer test result has a result with substantial overlap compared with the individual tests.	A positive painful arc test result and a positive external rotation resistance test result were the most accurate findings for detecting RCD.
Hanchard *et al*., 2013 Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement (Review)	Systematic Review	To evaluate the diagnostic accuracy of PETS impingements (subacromial or internal) or local lesions of bursa, rotator cuff or labrum that may accompany impingement, in people whose symptoms and/or history suggest any of these disorders.	Intervention: two pairs of review authors independently performed study selection, assessed the study quality using QUADAS.	The standard tests were Hawkins' test, Neer's sign and the painful arc test. The modified tests were Neer's sign, passive horizontal adduction, Speed's test and Yergason's test. There were three combination tests, which comprised: all of seven specific tests (drop arm test, Hawkins' test, Neer's sign, painful arc, passive horizontal adduction, Speed's test and Yergason's test); Hawkins' test or Neer's sign; and Hawkins' test and Neer's sign. The specificity estimates ranged from 26% for the standard Hawkins' test to 99% for the Gum-Turn test. Only one test was performed and interpreted similarly in two studies. This was the standard Hawkins' test, but different and possibly incomparable reference standards were used.	There is insufficient evidence upon which to base selection of PETS impingements, including subacromial impingement, in primary care.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Hegedus *et al.*, 2008 Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests	Systematic Review and meta-analysis	To compile and critique research on the diagnostic accuracy of individual orthopedic physical examination tests.	Intervention: the quality of the articles was assessed using the QUADAS tool and data was extracted from each article. Meta-analysis was performed on the Neer test and the Hawkins-Kennedy test for impingement.	Forty-five studies were critiqued with only half demonstrating acceptable high quality and only two having adequate sample size. For impingement, the meta- analysis revealed that the pooled sensitivity and specificity for the Neer test was 79% and 53%, respectively, and for the Hawkins-Kennedy test was 79% and 59%, respectively. Regarding orthopedic special tests where meta-analysis was not possible either due to lack of sufficient studies or heterogeneity between studies, no tests for impingement demonstrated significant diagnostic accuracy.	Based on pooled data, the diagnostic accuracy of the Neer test and the Hawkins-Kennedy test for impingement is limited. There is a great need for large, prospective, well-designed studies that examine the diagnostic accuracy of the numerous PETS.
Hegedus *et al.*, 2012 Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests	Systematic Review and meta-analysis	To update previously published systematic review and meta-analysis by subjecting the literature on shoulder physical examination to careful analysis in order to determine each tests clinical utility.	Intervention: this review is an update of previous work (see article n.6 in table). The QUADAS 2 tool was used to critique the quality of each new paper.	Since the publication of the 2008 review, 32 additional studies were identified and critiqued. For subacromial impingement, the meta-analysis revealed that the pooled sensitivity and specificity for the Neer test was 72% and 60%, respectively, for the Hawkins-Kennedy test was 79% and 59%, respectively, and for the painful arc was 53% and 76%, respectively.	Based on data from the original 2008 review and this update, the use of any single shoulder physical examination test to make a pathognomonic diagnosis cannot be unequivocally recommended. These findings seem to provide support for stressing a comprehensive clinical examination including history and physical examination

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Alqunaee *et al*., 2012 Diagnostic Accuracy of Clinical Tests for Subacromial Impingement Syndrome: A Systematic Review and Meta-Analysis	Systematic Review and meta-analysis	To examine the accuracy of clinical tests for diagnosing subacromial impingement syndrome (SIS).	Intervention: the methodological quality of selected studies was assessed using the QUADAS tool, a validated tool for the quality assessment of diagnostic accuracy studies. Pattern: authors included prospective or retrospective cohort studies that examined individuals with a painful shoulder, reported any clinical test for SIS, and used arthroscopy or open surgery as the reference standard.	The Hawkins-Kennedy test, Neer's sign, and empty can test are shown to be more useful for ruling out rather than ruling in SIS, with greater pooled sensitivity estimates (range, 0,69–0,78) than specificity (range, 0,57–0,62). A negative Neer's sign reduces the probability of SIS from 45% to 14%. The drop arm test and lift-off test have higher pooled specificities (range, 0,92–0,97) than sensitivities (range, 0,21–0,42), indicating that they are more useful for ruling in SIS if the test is positive.	Accurate diagnosis of SIS is a challenge to clinicians, and the diagnostic accuracy of some of the clinical tests used in clinical practice needs to be considered in the context of the overall patient assessment.
May *et al*., 2010 Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review	Systematic Review	To systematically review the reliability of physical examination procedures used in the clinical examination of patients with shoulder pain.	Intervention: pre-established criteria were used to judge the quality of the studies and satisfactory levels of reliability. A qualitative synthesis was performed based on levels of evidence.	Thirty-six studies were included with a mean methods score of 57%. Seventeen studies were deemed to be of high quality. The majority of studies indicated poor reliability for all procedures investigated.	There was conflicting evidence about their reliability, and most fell below the predetermined levels of acceptable reliability. Using these procedures to make their associated diagnoses is an invalid and unreproducible process.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Beaudreuil *et al.*, 2009 Contribution of clinical tests to the diagnosis of rotator cuff disease: A systematic literature review	Systematic Review	To evaluate the diagnostic performance of clinical tests for degenerative RCD, based on a systematic literature review.	Intervention: studies were selected based on the following criteria: patients with RCD, number of patients known to be greater than 30, known reference criterion (sub- acromial impingement or tendon disease), well-described diagnostic method, knowledge of the prevalence of the abnormality in the overall study population, well-described clinical tests with clear definitions of positive and negative test results, and availability of sensitivity and specificity data.	Authors selected nine studies, of which three investigated tests for subacromial impingement syndrome and seven tests for rotator cuff tendinopathy. The results fail to clarify the value of the painful arc test for diagnosing subacromial impingement syndrome. Two studies evaluated the Neer and Hawkins tests. Both tests were sensitive.Specificity, in contrast, was low. Few data are available on the Yocum test. Sensitivity was estimated at 78% versus arthroscopic evidence of tendinopathy with or without tearing.	The most extensively studied tests for subacromial impingement - Neer and Hawkins - are sensitive but lack specificity. Nevertheless, studies of the performance of clinical tests are few. The precise techniques and subjective interpretation required by clinical tests lead to substantial inter-observer variability.
Hughes *et al.*, 2008 Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review	Systematic Review	To evaluate the diagnostic accuracy of clinical tests to diagnose rotator cuff pathology.	Intervention: to reduce sources of bias, three reviewers independently assessed the included studies for methodological quality using criteria adapted from guidelines for appraising studies concerned with diagnostic tests by the National Health and Medical Research Council (1999). Data were extracted from each included study using a standard form developed for the review.	Thirteen studies met the inclusion criteria. The 13 studies evaluated 14 clinical tests. Most tests for rotator cuff pathology were inaccurate and cannot be recommended for clinical use.	The poor accuracy of clinical tests for rotator cuff pathology could be related to a lack of anatomical validity of the tests or it may be that the close relationships of structures in the shoulder may make it difficult to identify specific pathologies with clinical tests.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Dinnes *et al*., 2003 The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review	Systematic Review	The main objectives were as follows: 1. to establish the effectiveness of clinical examination and patient history in the differential diagnosis of shoulder pain; 2. to estimate the added benefit gained from use of diagnostic imaging for the identification of soft tissue shoulder disorders; 3. to assess how the individual tests could most effectively and cost-effectively be combined with clinical examination in diagnostic strategies or algorithms; 4. to identify gaps in the literature for the purpose of informing future research.	Intervention: authors proposed to appraise the methodological quality of included test accuracy studies using a formal quality assessment tool. Pattern: the primary inclusion criteria for the assessment of test accuracy were studies of clinical examination, ultrasound, MRI or MRA in patients suspected of having soft tissue shoulder disorders. Outcomes assessed were clinical impingement syndrome or rotator cuff tear (RCT) (full, partial or any). Only cohort studies were included.	The majority of studies evaluated the ability of clinical examination to identify patients with RCTs or, in two cases, specifically full-thickness RCTs. Two other studies used clinical examination to differentiate impingement syndrome from other causes of shoulder pain, and another evaluated a clinical test that aimed to distinguish outlet from non-outlet impingement syndrome, that is, those whose symptoms were not due to RCT or bursitis but were largely due to labral lesions or tears. The tests evaluated tended to be either highly sensitive or highly specific, and very few demonstrated both high sensitivity and specificity. As a result, few tests provided convincing evidence of the presence or absence of disease in the settings in which they were applied.	Too few studies evaluated the same test to draw any conclusions regarding individual tests. Insufficient evidence was found to recommend any one clinical examination test or set of tests or to provide an indication of the accuracy of clinical examination at differentiating RC disorders (as opposed to tears) from other causes of shoulder pain.

Table III. Synoptic table articles included for instrumental examinations.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
McGarvey *et al*. 2016 Diagnosis of rotator cuff tears using 3-Tesla MRI versus 3-Tesla MRA: a systematic review and meta-analysis	Systematic Review	To compare the diagnostic accuracy of MRI (3D) and MRA (2 & 3D) in the diagnosis of rotator cuff tears when performed exclusively at 3-T.	Intervention: a systematic review was undertaken of the Cochrane, MEDLINE and PubMed databases in accordance with the PRISMA guidelines. Methodological appraisal was performed using QUADAS 2. Pooled sensitivity and specificity were calculated and summary receiver-operating curves generated.	Both 3-T MRI and 3-T MRA showed similar excellent diagnostic accuracy for full-thickness supraspinatus tears. Concerning partial thickness supraspinatus tears, 3-T 2D MRA was significantly more sensitive (86.6 vs 80.5 %, p = 0.014) but significantly less specific (95.2 vs 100 %, p < 0.001). There was a trend towards greater accuracy in the diagnosis of subscapularis tears with 3-T MRA.	Three-Tesla MRI appeared equivalent to 3-T MRA in the diagnosis of full- and partial-thickness tears, although there was a trend towards greater accuracy in the diagnosis of subscapularis tears with 3-T MRA.
Roy J-S *et al*. 2015 Diagnostic accurancy of ultrasonography, MRI and MR arthrography in the characterisation of rotator cuff disorders: a systematic review and meta-analysis	Systematic Review and Meta-analysis	The objective was to perform a meta-analysis on the diagnostic accurancy of medical imaging for characterisation of RC disorders. A secondary analysis assessed accurancy of US by radiologists and non- radiologists.	Intervention: a systematic search in three databases was conducted. Two raters performed data extraction and evaluation of risk of bias independently (QUADAS 2), and agreement was achieved by consensus.	Diagnostic accuracy of US, MRI and MRA in the characterisation of full-thickness RC tears was high with overall estimates of sensitivity and specificity over 0.90. As for partial RC tears and tendinopathy, overall estimates of specificity were also high (> 0.90), while sensitivity was lower (0.67-0.83). Diagnostic accuracy of US was similar whether a trained radiologist, sonographer or orthopedist performed it.	Our results show the diagnostic accuracy of US, MRI and MRA in the characterisation of full thickness RC tears. Regarding the partial cuff lesions and tendinopathy the specificity is high, the sensitivity lower. When considering accuracy, cost, and safety, US is the best option.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Lenza M. *et al.* 2013 Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered	Systematic Review	To compare the diagnostic test accuracy of MRI, MRA and US for detecting any rotator cuff tears (i.e., partial or full thickness) in people with suspected rotator cuff tears for whom surgery is being considered.	Intervention: the research was carried out on various databases and the quality assessment was performed according to the QUADAS criteria.	We included 20 studies of people with suspected rotator cuff tears, of which six evaluated MRI and US, or MRA and US in the same people.	MRI, MRA and US have good diagnostic accuracy and any of these tests could equally be used for detection of full thickness tears in people with shoulder pain for whom surgery is being considered. The diagnostic performance of MRI and US may be similar for detection of any rotator cuff tears. However, both MRI and US may have poor sensitivity for detecting partial thickness tears, and the sensitivity of US may be much lower than that of MRI.
Amee Seitz *et al.* 2010 Ultrasonographic Measures of Subacromial Space in Patients with Rotator Cuff Disease: A systematic Review	Systematic Review	Evaluate through the use of US how much the problems of rotator cuff can modify the measure of the subacromial space.	Intervention: the search was carried out on various databases (CENTRAL, PubMed, PEDro, Medline, CIHAHL Plus and SPORTDiscus).	5 studies were included.	US showed a lower subacromial distance (AHD) in patients with complete cuff injuries compared to healthy subjects and with subacromial impingement.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Ottenheijm *et al*. 2010 Accurancy of Diagnostic Ultrasound in Patients With Suspected ubacromial Disorders: A Systematic Review and Meta-Analysis	Systematic Review and Meta-analysis	To determine diagnostic accuracy of US for detecting subacromial disorders in patients presenting in primary and secondary care settings	Intervention: the search was carried out on Medline and Embase.	23 studies were included. Below are the sensitivity and specificity values of ultrasound for each type of subacromial disorder considered. Complete lesions (sensitivity 0.95, specificity 0.96). Partial lesions (sensitivity 0.72, specificity 0.93). Subacromial bursitis (sensitivity from 0.79 to 0.81 and specificity from 0.94 to 0.98); Tendinopathy (sensitivity from 0.67 to 0.93 and specificity from 0.88 to 1.00); Tendon calcifications (specificity 1.00 and specificity from 0.88 to 0.98).	US is recommended in those patients in whom conservative treatment fails to detect or exclude total lesions, partial lesions and to a lesser extent, to diagnose tendinopathy, subacromial bursitis and calcific tendinitis.
De Jesus J.O. *et al*. 2009 Accurancy of MRI, MR Arthrography and Ultrasound in the diagnosis of rotator cuff tears: a meta-analysis	Meta-analysis	The purpose of this study was to compare the diagnostic accuracy of MRI, MRA and US for the diagnosis of rotator cuff tears through a meta-analysis of the studies in the literature.	Intervention: a Medline search was performed.	65 articles were included. In the diagnosis of complete or partial lesions, MRA is more sensitive and specific than MRI and ultrasound.	MRA is the most sensitive and specific technique for detecting full-thickness and / or partial thickness cuff injuries. US and MRI have similar sensitivity and specificity.

Article Author and publication date	Study design	Objective	Methods	Results	Conclusions
Dinnes J *et al*. 2003 The effectiveness of diagnostic test for the assessment of shoulder pain due to soft tissue disorders: a systematic review	Systematic Review	Evaluate the efficacy and cost / benefit ratios of the diagnostic imaging methods, in addition to the clinical examination and the patient's medical history, in the diagnosis of shoulder soft tissue disorders. To establish the effectiveness of the clinical examination and the patient's medical history in the differential diagnosis of shoulder pain; Estimate the added benefit deriving from the use of diagnostic imaging in the identification of shoulder soft tissue disorders; Evaluate how individual tests can be combined more effectively and economically with clinical examination in diagnostic strategies or algorithms	Intervention: the literature has been investigated from various sources. The methodological quality of the included studies was carried out using a formal assessment tool for the quality of diagnostic studies and data extraction was performed in duplicate. For each imaging test, positive and negative sensitivity, specificity and LR were calculated with 95% confidence intervals for each study.	10 cohort studies were included for the clinical examination, 38 for ultrasound, 29 for MRI and 6 for MRA. For the clinical examination the identified studies evaluated the accuracy of the individual tests or the clinical examination in its entirety: the individual tests had high sensitivity and specificity, but the small sample size did not allow conclusions to be drawn. US showed greater accuracy in detecting full-thickness lesions while showing lower sensitivity in the detection of partial lesions. MRI showed high sensitivity and specificity in the detection of total lesions, but a lower sensitivity for partial lesions. MRA has proved to be very accurate in detecting total lesions, but less accurate in detecting partial lesions.	Clinical examination by specialists can exclude the presence of a cuff lesion and both MRI and ultrasound can be used in the same way to detect full-thickness lesions, although ultrasound may be better in detecting lesions partial. Furthermore, ultrasound may also be more convenient for the detection of total lesions in a specialised hospital setting.

Data extraction

Study design, objectives, methods, results, and conclusions have been extracted from the included studies (see **tables II, III**). Methodological quality has been assessed through the AMSTAR 2 systematic review checklist (23) (**appendix 5**).

Synopses of the results

Studies on diagnostic psychometric quality of clinical tests

Gismervik *et al.* (2017) (18) performed a meta-analysis and analysed 20 articles with a low risk of bias according to the QUADAS score (Quality Assessment of Diagnostic Accuracy Studies), with the aim of using the diagnostic odds ratio (DOR) to evaluate overall efficacy of clinical shoulder tests. Specifically, the tests analysed for this SIS were the Neer and Hawkins-Kennedy tests. From this analysis the Hawkins-Kennedy test obtained the highest DOR (2.86) with a sensitivity of 58% and a specificity of 67%, unlike the Neer test which obtained a DOR of 2.17 with a sensitivity of 59% and a specificity of 60%. However, it appears that no single test has shown higher clinical efficacy than another (18).

Lange *et al.* (2017) (24) included 18 prospective studies that examined the reliability of 62 clinical tests for the diagnosis of shoulder disorders; 17 of these studies examined the inter-rater reliability, while only one study assessed the intra-rater reliability. The meta-analysis identified a wide heterogeneity for the Hawkins-Kennedy test, Neer Test, Empty Can Test, Painful Arc Test with a statistical I2 index > 0.75 which can be interpreted as "considerable heterogeneity" according to the Cochrane manual (24, 25). The results of the meta-analysis indicate an inter-examiner reliability from moderate to substantial in the Hawkins-Kennedy Test, Neer Test, Empty Can Test and Painful Arc Test. This study recommends interpreting the results with caution since they have limited reliability and poor validity (24).

O'Kane *et al.* (2014) (26) analysed 32 articles and identified 57 clinical tests for the shoulder. Specifically, the Hawkins-Kennedy test and the Neer test were analysed through the results of Hegedus's meta-analysis included in the study. This reported high sensitivity (80%) and low specificity (56%) for the Hawkins-Kennedy test. The authors maintain that low specificity reduces the diagnostic accuracy of the test with a positive likelihood ratio (LR +) of 1.84 and a negative likelihood ratio (LR -) of 0.35. Similar results were obtained for the Neer test, with an LR + of 1.79 and LR- of 0.47. Thus, for both tests, a positive result slightly increases the probability of disease, and a negative result slightly decreases this probability. Therefore, used individually, clinical tests for SIS have a low diagnostic capacity (26).

Hermans *et al.* (2013) (27) performed a meta-analysis on 28 studies which evaluated shoulder clinical examination. Among the pain provocation tests, the positivity at painful arc test was the only result with an LR + greater than 2.0 and the negative painful arc had the lowest negative LR. The positive results of the Hawkins test (LR 1.5) and the Neer test (range LR 0.98 - 1.6) was of little value. Because of low diagnostic accuracy of the individual tests performed, multiple test combinations were evaluated. However, a positive result of the combination of the Neer test and the Hawkins test (LR 1.6) overlaps with that of the tests performed individually. So even the combination of these tests does not improve their diagnostic accuracy (27).

Hanchard *et al.* (2013) (4) included 33 studies. 4 of the studies included, analysed 13 tests for SIS (standard, modified or combinations of tests) whose specificity and sensitivity were evaluated. Standard tests: Hawkins test (specificity 0.26-0.44; sensitivity 0.92), Neer sign (specificity 0.32; sensitivity 0.89) and painful arc (specificity 0.82; sensitivity 0.32). Modified tests: Neer sign in internal rotation (specificity 0.48; sensitivity 0.75), passive horizontal adduction (specificity 0.29; sensitivity 0.82), Speed's test (specificity 0.55; sensitivity 0.69) and Yergason test (specificity 0.87, sensitivity 0.37). Finally, three test combinations were analysed: seven specific tests (drop arm test, Hawkins test, Neer sign, painful arc, passive horizontal adduction, Speed's test and Yergason test) (specificity 0.97; sensibility 0.05); Hawkins test or modified Neer sign (specificity 0.41; sensibility 0.96); Hawkins test and modified Neer sign (specificity 0.51; sensibility 0.71). One study evaluated the internal rotation resistance strength test to differentiate SIS from internal impingement and a sensitivity of 88% and a specificity of 96% were highlighted.

However, by collected data, the authors conclude that there is not enough evidence on which to base the selection of clinical tests for SIS in a primary care setting. Moreover, an extreme diversity in performance and in the interpretation of the tests has been detected, which hinders the synthesis of evidence and/or clinical applicability (4).

Hegedus *et al.* (28) elaborated two systematic reviews with meta-analysis of which the second is an update of the first, in order to determine the clinical utility of each test. In the first review 45 studies were included; however, the meta-analysis was only possible for 3 tests. Regarding SIS, the tests analysed in the meta-analysis were the Hawkins and the Neer test. The analysis showed the sensitivity and specificity of the Neer test are 79% and 53% respectively, while for the Hawkins test they are 79% and 59%. However, statistical

results suggested that these tests were not significant in diagnostic utility for SIS (29). In the second revision of 2012, a further 32 articles were included; the results had little variations (Neer test: sensitivity 72%, specificity 60%; Hawkins test: sensitivity 79%, specificity 59%); In the meta-analysis of the test for SIS it was also included the painful arch with a sensitivity of 53% and a specificity of 76%. However, the diagnostic odds ratio (DOR) for any of the 3 tests examined indicates that such tests, when positive, are unlikely to have the diagnostic capacity to discriminate patients with SIS (28).
Alqunaee *et al.* (2012) (30) performed a systematic review including 16 prospective and retrospective cohort studies, of which 10 were included in the meta-analysis. The results of the study show that the Hawkins-Kennedy test, the Neer sign and the empty can test are more useful for excluding rather than confirming SIS. The negative Neer sign reduces the probability of the presence of SIS from 45% to 14%. Instead, the lift-off test provides clear evidence to confirm SIS when the test is positive (LR + 16.47) (30).
May *et al.* (2010) (31) included 36 studies about shoulder physical examination procedures. The results were grouped according to the anatomical or pathological entities examined or based on certain procedures of the physical examination. For all the procedures contradictory results emerged. The inter-rater reliability from test that assess SIS was evaluated in 6 studies. Two of these studies showed for some tests (painful arch, Hawkins-Kennedy test, Neer sign) values of the kappa coefficient > 0.85, suggesting "almost perfect" concordance. The results for intra-rater reliability showed kappa coefficient equal to 1 -in one study-, suggesting "almost perfect" agreement.
The study shows there is conflicting evidence on the reliability of physical examination procedures for the evaluation of shoulder pain and the majority has a reliability below the predetermined acceptable levels (31).
Beaudreuil *et al.* (2009) (32) selected 9 studies, of which 3 examined the tests for SIS.
The painful arc during lateral elevation of the arm was evaluated in two studies that produced conflicting results in terms of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). In addition, the accuracy of this test was poor. So, the results fail to clarify the value of the painful arc test for the diagnosis of SIS.
Two studies evaluated the Neer test and the Hawkins test. Both tests were very sensitive (Neer, 75-89%; and Hawkins, 91-92%). Specificity, instead, was low (Neer, 30-40%; Hawkins, 25-44%). PPV and NPV differed between the two studies. Diagnostic accuracy was determined in one study and showed that 72/75% of cases were correctly classified by both tests. The combination of the two tests produced similar performance characteristics.
Sensitivity of Yocum test was estimated at 78%.
This study shows that the most studied tests for SIS - Neer and Hawkins - are sensitive but lack specificity. So, the precise techniques and subjective interpretation, lead to substantial inter-observer variability (32).
Hughes *et al.* (2008) (20) produced a systematic review including 13 studies that included 89 diagnostic accuracy assessments that were determined using the likelihood ratios (LRs) of 14 clinical tests. Regarding SIS, the results show that the Hawkins-Kennedy test was the only one to have produced a positive likelihood ratio greater than 10 or a negative likelihood ratio of less than 0.1. However, this result was not found in 6 other evaluations carried out in 3 studies. The Neer test and the horizontal adduction test were shown to be inaccurate for the diagnosis of SIS (20).
Dinnes *et al.* (2003) (33) included in their review 10 cohort studies. 2 of these studies used the clinical examination to differentiate the SIS from other causes of shoulder pain and one study evaluated Internal rotation resistance strength test. Evaluated tests tended to be highly sensitive or highly specific and very few demonstrated both high sensitivity and specificity. Consequently, few tests have provided convincing evidence of the presence or absence of pathology in the setting in which they were applied. Several tests have produced a sensitivity > 80% for the detection of SIS, including the Neer test, the Hawkins test, the horizontal adduction test, the Jobe test, the impingement sign, and the painful arc. Tests with specificities > 80% included the drop arm test, Yergason test, Speed test and external passive rotation. Notably, the small size of the sample examined did not provide sufficient evidence to prove that these tests are really accurate. About the combination of two or more tests, the negative likelihood ratios proved to be sufficiently low to confirm that the pathology is absent in those with a negative diagnosis (33).

Studies on diagnostic accuracy of instrumental examinations

In Mc Garvey *et al.*' s study (2016) (34) 14 articles were included that investigated the diagnostic accuracy of MRI and MRA in 2D and 3D (index test) compared to arthroscopic surgery on rotator cuff lesions. The sensitivity and specificity of MRI and MRA were over 95% for full thickness supraspinatus tears. For partial thickness supraspinatus tears, the results concerning sensitivity and specificity showed great variability based on the type of imaging used (Sensitivity: MRI 80.5, MRA 2D 86.5, MRA 3D 82.7; Specificity: MRI 100, MRA 2D 95.2, MRA 3D 93.9) as well as for the lesions of the subscapularis (Sensitivity: MRI 64.2, MRA 2D 80.9, MRA 3D 83.1; Specificity: MRI nd, MRA 2D 86.3, MRA 3D 82.0) (44).

Roy *et al.*'s study (2015) (17) compared the diagnostic accuracy of US, MRI and MRA in the diagnosis of lesions of the rotator cuff (full thickness and partial thickness). Furthermore, secondary analyses were carried out according to the technical characteristics of the imaging equipment (dividing studies that included US < or equal to 7.5 MHz and with MRI and MRA < or equal to 1.5 T from studies with US > 7.5 MHz and MRI - MRA at 3.0 T). All types of imaging considered reported a similar performance in the detection of rotator cuff lesions of any type (Sensitivity: US 0.91, MRI 0.90, MRA 0.90; Specificity: US 0.86, MRI 0.90, MRA 0.90). For full thickness tears (Sensitivity: US 0.91, MRI 0.90, MRA 0.90; Specificity: US 0.93, MRI 0.93, MRA 0.95) and partial thickness tears, the results are similar even if the sensitivity of US and MRI for partial thickness tears was lower (US 0.68 and MRI 0.67), in the arthro-resonance the sensitivity was higher (0.83), while the specificity reported was equivalent for the three diagnostic modalities (US 0.94, MRI 0.94, MRA 0.93) (17).

The goal of the Cochrane Review by Lenza *et al.* (2013) (16) was to investigate the diagnostic accuracy of US, MRI and MRA in subjects with suspected lesion of the rotator cuff tendons for which surgery was considered. 272 articles were identified for systematic review and 82 for meta-analysis. Studies that used MRI, MRA and US as an index test and surgery (arthroscopic or open surgery) as a reference standard were included. No statistically significant differences were found for the sensitivity and specificity between US and MRI globally in the detection of full-thickness tears of the rotator cuff regardless of size (sensitivity: MRI = 98%, US = 91%; specificity: MRI = 79 %, US = 85%), as well as no differences between US, MRI and MRA were found in the detection of full-thickness rotator cuff tears with sensitivity and specificity of 92% or more (sensitivity: MRI = 94%, MRA = 94%, US = 92%; specificity: MRI = 93%, MRA = 92%, US = 93%). Instead, about the partial thickness tears, US and MRI revealed a lower sensitivity (MRI = 74%, US = 52%), while the specificity was high (MRI and US = 93%). The specificity of all three imaging techniques was good, except in the case of detection of rotator cuff lesions in general, with no distinction according to size (16).

In the review by Seitz *et al.* (2011) (35) 5 case-control studies were included in which the Acromion Humeral Distance (AHD) was measured by US, both on asymptomatic subjects and on subjects with rotator cuff disorders. It was found that the AHD measured with the patient's arm along the side (arm by side) was lower in subjects with full-thickness rotator cuff lesions than in healthy subjects and with SIS (asymptomatic: 10.5 mm; tendinopathy: 14.4 mm; partial rotator cuff tears: 10.8 mm; total rotator cuff tears: 8.6 mm). In addition, lower AHD was detected in subjects with SIS than in healthy subjects during active arm elevation (35).

In the study by Ottenheijm *et al.* (2010) (36), the diagnostic accuracy of the US used to detect SIS was assessed. In the 23 studies investigating the US scans the sensitivity and specificity results for each type of subacromial disorder have been as follows: complete lesions (sensitivity 0.95, specificity 0.96), partial lesions (sensitivity 0.72, specificity 0.93), subacromial bursitis (sensitivity from 0.79 to 0.81 and specificity from 0.94 to 0.98), tendinopathy (sensitivity from 0.67 to 0.93 and specificity from 0.88 to 1.00) (36).

The study by De Jesus *et al.* (2009) (37) compared the diagnostic accuracy of MRI, MRA and US in the diagnosis of rotator cuff lesions through a meta-analysis. 65 articles were included and the results of sensitivity and specificity are as follows: full thickness tears (Sensitivity: MRA 95.4; MRI 92.1; US 92.3; Specificity: MRA 98.9; MRI 92.9; US 94.4), partial thickness tears (Sensitivity: MRA 85.9; MRI 63.6; US 66.7; Specificity: MRA 96.0; MRI 91.7; US 93.5), total or partial thickness tears (Sensitivity: MRA 92.3; MRI 87.0; US 85.1; Specificity: MRA 94.5; MRI 81.7; US 86.1) (37).

Dinnes *et al.* (2003) (33) compared different imaging techniques (MRA, MRI and US) with the aim of assessing their accuracy and cost-effectiveness in the diagnosis of shoulder soft tissue disorders in addition to the clinical examination and patient' s history. 38 studies were included for US, including 19 on the detection of full-thickness rotator cuff tears (Sensitivity: 0.87; Specificity 0.96) and 11 for partial thickness tears (Sensitivity: 0.67; Specificity: 0.94); for MRI 29 studies were included and sensitivity and specificity values were reported for the detection of full thickness tears (Sensitivity: 0.89; Specificity 0.93), partial thickness tears (Sensitivity 0.44; Specificity 0.90) and all types of lesions (Sensitivity: 0.83; Specificity: 0.86). Finally, for the MRA, 6 studies were included which revealed the following sensitivity and specificity values for full thickness tears (Sensitivity: 0.95; Specificity: 0.93), partial thickness tears (Sensitivity: 0.62; Specificity: 0.92) and for all the types of lesions (Sensitivity: 0.88; Specificity: 0. 83) (33).

DISCUSSION

The objectives of this systematic review were 1) to investigate the utility of clinical tests and their psychometric qualities, and 2) to investigate the accuracy of different imaging methods in the clinical diagnosis of SIS.

Psychometric qualities of clinical tests

The first question we tried to answer through this review is if psychometric qualities of clinical tests were accurate for the diagnosis of SIS.

In this regard, several studies examined the likelihood ratios of clinical tests that aim to diagnose SIS and they agree that no test improves the post-test probability of detecting this type of conflict and, consequently, they are inaccurate (20, 27 30). This lack of post-test accuracy, may be due to the excessive emphasis on a diagnosis based on the recognition of the tissue involved and the underlying pathology that provides inconsistent information compared to those deriving from the analysis of signs and symptoms that can be referred to the pain (20).
According to the definition of the IASP (International Association for the Study of Pain), pain is "an unpleasant sensory and emotional experience associated with tissue damage, present or potential, or described in terms of damage" (38). Therefore, pain is experienced both as a physical sensation and as an emotional experience, perceived on a cognitive level as an intense and unpleasant sensation of discomfort which leads to a reactive behaviour; so, an analysis focused only on structural discrimination is limited and demonstrates critical issues.
In addition, the subacromial bursa (SAB) is often implicated in shoulder pain. The SAB is a richly innervated structure and is involved and stimulated by every movement of the arm; consequently, when it is reactive, it could create false positives in clinical maneuvers which reveal inaccurate in determining the pain generator (39).
This hypothesis is confirmed by several studies which have underlined a good sensitivity but a poor specificity of these tests which are, therefore, able to identify the population with shoulder pain, but not the specific structure "pain generator" (20, 27, 32, 33).
About the reliability of orthopedic clinical tests, a fundamental aspect is the inter-examiner concordance that is in various studies (4, 23, 30). The best evaluation for concordance is the kappa coefficient[4] and the values vary from 0 to 1 (40).
On the basis of these criteria, the authors of the various studies agree that there is a moderate agreement for the Painful Arc test, while for the Neer test the concordance ranges from modest to moderate, for the Empty Can test and the Hawkins test they have found only a modest concordance. Therefore, none of these tests demonstrates a good or excellent inter-examiner agreement and consequently an acceptable level of reliability (31).
In addition, in the systematic review of Hanchard *et al.* (4) has been evidenced that the majority of the literature has revealed an extreme diversity both in the performance and in the interpretation of the tests that hinders their evidence and/or uniform clinical applicability (4).

Diagnostic accuracy of instrumental examination

To evaluate the accuracy of US, MRI and MRA, the indirect signs of SIS (*i.e.*, structural alterations of rotator cuff tendons) have been considered; notably, sensitivity and specificity in detection of partial and full rotator cuff injuries have been investigated.
From the studies analyzed, some results have emerged on which it is necessary do some considerations: US, MRI and MRA have the same level of sensitivity and specificity in the detection of complete lesions of the rotator cuff (16, 17, 33). the US has the best cost-effectiveness ratio of all three diagnostic methods even if it is administered by non-radiologist professionals, while MRA appears to be more precise for subscapularis lesions and is the most invasive (17, 33, 34, 37). For partial injuries it seems that there is a lower sensitivity in view of high specificity, especially the US (16, 17, 36).
In the light of the results of this study, it appears that the diagnostic system we have available to screen the patient with SIS is methodologically poor as well as pragmatically inadequate; this could be caused by the persistence of clinicians to give diagnostic labels which, as recent literature has shown, lack uniformity, accuracy and reliability (41).
In this regard, Hegedus *et al.*, defined the term "impingement" all-encompassing for the pathology of rotator cuff tendons that at best, is little useful in the guide to treatment; in the worst is, a clinical illusion; this diagnostic label does not provide information on diagnosis, treatment or prognosis (31).
So recent literature is oriented to consider the diagnostic label of SIS as a clinical hypothesis rather than a diagnosis based on the correlation between structural alterations and symptoms, so shoulder surgeons begin to replace the term "subacromial impingement" with the most realistic "anterior shoulder pain" (8).
This nomenclature, therefore, does not represent a diagnostic entity detectable with specific structural tests or imaging techniques, but a functional clinical entity (4, 13).
To prove the weak correlation between structure and pain in SIS, numerous studies have not found significantly different values between real and placebo surgery: the "conflict to be operated" was not the cause of the problem (7, 42).
In light of this work, the Cochraine revision of Karjalainen, 2019, concludes that RCT's High Evidence showed that Subacromial decompression does not provide clinically important benefits over placebo in pain, function or quality of life (43).
Rotator cuff injuries are just some of the structural alterations that have been associated with the presence of SIS (11). However, the diagnosis of SIS cannot be based exclusively on their presence since in the rotator cuff tendons structural anomalies are present in 20% of the general population and of this 20%, 2/3 are asymptomatic (14-16). The presence of structural alterations in a large number of asymptomatic subjects, challenges the ability of the imaging to identify such alterations as "pain generators" (12);

therefore, all clinicians should not blindly apply the results of bio-imaging.
Results of this review showed that clinical tests have shown a limited ability to make diagnosis of SIS; for this reason they must be interpreted with great caution in terms of diagnostic value and decision-making process since they have a limited reliability and a poor validity both if performed singularly and in combination (4, 18, 20, 24, 26-28, 31, 32). Furthermore, the majority of the clinical tests for SIS are inaccurate and so it cannot be recommended for clinical use (4, 18, 20, 24, 26, 27, 31, 32). The only clinical transferability they can have is connected to adequate therapeutic communication with another professional.
However, in current clinical practice, patients are still screened to undergo surgery based on the answers of clinical tests and diagnostic imaging; while recent research suggests that fewer surgical procedures for SIS should be performed; surgical treatment should be considered only after the failure of evidence based conservative care for at least 3 months, based on the patient's education in pain and exercise (12).
Lewis 2009 (45) and Littlewood 2012 (44) propose to replace the no more descriptive diagnostic label of the underlying pathological process with terms like "Rotator cuff related shoulder pain" / "tendinopathy" because they enable a well-defined management strategy and prognostic profile to reassure the patient mainly, despite the lack of a precise structural diagnosis, emphasizing the acceptability and feasibility of a conservative treatment as the first useful and concrete response in case of painful shoulder (43). The management of painful shoulder therefore includes the patient's reassurance and education, the gradual exposure to the load by therapeutic exercise for a program lasting no more than 12 weeks (44).
The structural view of shoulder pain rests on a weak diagnostic and confirmatory process, in which the patient with a painful shoulder and an associated lesion independent of the ongoing symptomatology could be directed to surgery (46). Through the years, an increase of the rate of surgery for SIS has been a consequence of this deficient diagnostic system and of the resulting and inaccurate diagnostic labels (47) even if literature highlights same results between conservative treatment and surgery in terms of pain reduction and quality of live - until 10-year follow-up - with the disadvantage that surgery is more expensive and exposes the patient at a greater risk of complications, for which conservative treatment is the best choice (48).
Results of this review show that clinicians should leave the diagnostic pathology model based on tests that lack validity and reliability and poor instrumental diagnostic examinations because it has been shown that pain is not always dependent rom structural failures (14, 31, 49-51).

Points of strength and limits

The strengths of this review are the presence of several reviewers and the inclusion of only systematic reviews and/or meta-analysis.
Unfortunately, most of the included studies demonstrated a poor methodological quality according to the AMSTAR 2 **checklist** (23). Furthermore, the included studies are only in English and no knowledge of other relevant studies in other languages is known.
Another main limitation of this revision is the research of the studies through a single database.

CONCLUSIONS

The present review confirms the poor diagnostic ability of clinical tests for SIS since the poor values of diagnostic accuracy and excessive heterogeneity of performance do not allow an adequate applicability in clinical practice. Furthermore, this review highlights the criticism in imaging employment in the complex process of the clinical framing of SIS.
Diagnostic labeling based on a pathological anatomical classification not always reflect the clinical path, and would be advisable to manage the patient with SIS with procedures leave no longer focused on special tests and confirmatory instrumental examinations, but rather on the framing of bio-psycho-social factors for optimal and taylorised management.

CONFLICT OF INTERESTS

The authors declare that they have no conflict of interests.

REFERENCES

1. Luime J, Koes B, Hendriksen I. Prevalence and incidence of shoulder pain in the general population: a systematic review. Scand J Rheumatol 2004;33:73-81.
2. Bot SD, Van der Waal JM, Terwee CB. Incidence and prevalence of complaints of the neck and upper extremity in general practice. Ann Rheum Dis 2005;64:118-23.

3. Tempelhof S, Rupp S, Seil R. Age-related prevalence of rotator cuff tears in asymptomatic shoulders. J Shoulder Elbow Surg 1999;8:296-9.
4. Hanchard NC, Lenza M, Handoll HH, Takwoingi Y. Physical tests for shoulder impingements and local lesions of bursa, tendon or labrum that may accompany impingement. Cochrane Database Syst Rev 2013;2013(4):CD007427.
5. Östör AJK, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. Rheumatology 2005;44:800-5.
6. Michener LA, Walsworth MK, Burnet EN. Effectiveness of rehabilitation for patients with subacromial impingement syndrome: a systematic review. J Hand Ther 2004;17(2):152-64.
7. Neer CS. Anterior acromioplasty for the chronic impingement syndrome in the shoulder: a preliminary report. J Bone Joint Surg Am 1972;54:41-50.
8. Braman JP, Zhao KD, Lawrence RL, Harrison AK, Ludewig PM. Shoulder impingement revisited: evolution of diagnostic understanding in orthopedic surgery and physical therapy. Med Biol Eng Comput 2014;52:211-9.
9. Beltran LS, Nikac V, Beltran J. Internal Impingement Syndromes. Magn Reson Imaging Clin N Am 2012;20:201-11.
10. Gerber C, Sebesta A. Impingement of the deep surface of the subscapularis tendon and the reflection pulley on the anterosuperior glenoid rim: a preliminary report. J Shoulder Elbow Surg 2000;9:483-90.
11. Escamilla RF, Hooks TR, Wilk KE. Optimal management of shoulder impingement syndrome. Open Access J Sports Med 2014.
12. Lewis J, McCreesh K, Roy J-S, Ginn K. Rotator Cuff Tendinopathy: Navigating the Diagnosis-Management Conundrum. J Orthop Sports Phys Ther 2015;45:923-37.
13. McFarland EG, Maffulli N, Del Buono A, Murrell GA, Garzon-Muvdi JA, Petersen SA. Impingement is not impingement: the case for calling it "Rotator Cuff Disease". Muscles Ligaments Tendons J 2013;3(3):196-200.
14. Garving C, Jakob S, Bauer I, Nadjar R, Brunner UH. Impingement syndrome of the shoulder. Dtsch Arztebl Int 2017;114:765-76.
15. Ristori D, Miele S, Rossettini G, Monaldi E, Arceri D, Testa M. Towards an integrated clinical framework for patient with shoulder pain. Arch Physiother 2018,30;8:7.
16. Lenza M, Buchbinder R, Takwoingi Y, Johnston RV, Hanchard NC, Faloppa F. Magnetic resonance imaging, magnetic resonance arthrography and ultrasonography for assessing rotator cuff tears in people with shoulder pain for whom surgery is being considered. Cochrane Database Syst Rev 2013;2013(9):CD009020.
17. Roy J-S, Braën C, Leblond J. Diagnostic accuracy of ultrasonography, MRI and MRarthrography in the characterisation of rotator cuff disorders: a systematic review and meta-analysis. Br J Sports Med 2015;49:1316-28.
18. Gismervik SØ, Drogset JO, Granviken F, Rø M, Leivseth G. Physical examination tests of the shoulder: a systematic review and meta-analysis of diagnostic test performance. BMC Musculoskelet Disord 2017;25;18(1):41.
19. Papadonikolakis A, McKenna M, Warme W, Martin BI, Matsen FA. Published evidence relevant to the diagnosis of impingement syndrome of the shoulder. J Bone Joint Surg Am 2011;93:1827-32.
20. Hughes PC, Taylor NF, Green RA. Most clinical tests cannot accurately diagnose rotator cuff pathology: a systematic review. Aust J Physiother 2008;54:159-70.
21. Padulo J, Oliva F, Frizzero A, Maffulli N. Basic principles and recommendations in clinical and field Science Research: 2018 update. MLTJ 2018;8(3):305-7.
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Linee guida per il reporting di revisioni sistematiche e meta-analisi: il PRISMA Statement. OPEN ACCESS 2015;7:8.
23. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017:j4008.
24. Lange T, Matthijs O, Jain NB, Schmitt J, Lützner J, Kopkow C. Reliability of specific physical examination tests for the diagnosis of shoulder pathologies: a systematic review and meta-analysis. Br J Sports Med 2017;51:511-8.
25. Cochrane Handbook for Systematic Reviews of Interventions n.d. http://handbook-5-1.cochrane.org/index.htm#chapter_9/9_analysing_data_and_undertaking_meta_analyses.htm. Last access date: 07/21/2019.
26. O'Kane JW, Toresdahl BG. The Evidenced-Based Shoulder Evaluation: Curr Sports Med Rep 2014;13:307-13.
27. Hermans J, Luime JJ, Meuffels DE, Reijman M, Simel DL, Bierma-Zeinstra SMA. Does This Patient With Shoulder Pain Have Rotator Cuff Disease?: The Rational Clinical Examination Systematic Review. JAMA 2013;310:837.
28. Hegedus EJ, Goode AP, Cook CE. Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. Br J Sports Med 2012;46:964-78.
29. Hegedus EJ, Goode A, Campbell S. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br J Sports Med 2007.
30. Alqunaee M, Galvin R, Fahey T. Diagnostic Accuracy of Clinical Tests for Subacromial Impingement Syndrome: A Systematic Review and Meta-Analysis. Arch Phys Med Rehabil 2012;93:229-36.
31. May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy 2010;96:179-90.
32. Beaudreuil J, Nizard R, Thomas T. Contribution of clinical tests to the diagnosis of rotator cuff disease: A systematic literature review. Joint Bone Spine 2009;76:15-9.
33. Dinnes J, Loveman E, McIntyre L, Waugh N. The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review. Health Technol Assess 2003;7(29):iii, 1-166.
34. McGarvey C, Harb Z, Smith C, Houghton R, Corbett S, Ajuied A. Diagnosis of rotator cuff tears using 3-Tesla MRI versus 3-Tesla MRA: a systematic review and meta-analysis. Skeletal Radiol 2016;45:251-61.
35. Seitz AL, Michener LA. Ultrasonographic Measures of Subacromial Space in Patients with Rotator Cuff Disease: A Systematic Review. J Clin Ultrasound 2011;39(3):146-54.

36. Ottenheijm RP, Jansen MJ, Staal JB, *et al*. Accuracy of diagnostic ultrasound in patients with suspected subacromial disorders: a systematic review and meta-analysis. Arch Phys Med Rehabil 2010;91:1616-25.
37. De Jesus, Laurence Parker, Andrea J. Frangos. Levon N. Nazarian Accuracy of MRI, MR Arthrography, and Ultrasound in the Diagnosis of Rotator Cuff Tears: A Meta-Analysis. AJR Am J Roentgenol 2009;192(6):1701-7.
38. IASP Terminology-IASP. Available from: https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698#Pain. Last access date 09/25/2019.
39. Dean BJ, Gwilym SE, Carr AJ. Why does my shoulder hurt? A review of the neuroanatomical and biochemical basis of shoulder pain. Br J Sports Med 2013;47(17):1095-104.
40. Ashby D. Practical statistics for medical research. Douglas G. Altman, Chapman and Hall, London, 1991. No. of pages: 611. Price: £32.00. Stat Med 1991;10:1635-6.
41. Schellingerhout JM, Verhagen AP, Thomas S, Koes BW. Lack of uniformity in diagnostic labeling of shoulder pain: Time for a different approach. Man Ther 2008;13:478-83.
42. Lähdeoja T, Karjalainen T, Jokihaara J. Subacromial decompression surgery for adults with shoulder pain: a systematic review with meta-analysis. Br J Sports Med 2020;54(11):665-73.
43. Karjalainen TV, Jain NB, Page CM. Subacromial decompression surgery for rotator cuff disease.Cochrane Database Syst Rev 2019;1:CD005619.
44. Littlewood C. Contractile dysfunction of the shoulder (rotator cuff tendinopathy): an overview. J Man Manip Ther 2012;20:209-13.
45. Lewis JS. Rotator cuff tendinopathy/subacromial impingement syndrome: is it time for a new method of assessment? Br J Sports Med 2009;43:259-64.
46. Lewis J. The End of an Era? J Orthop Sports Phys Ther 2018;48(3):127-9.
47. Vitale MA, Arons RR, Hurwitz S, Ahmad CS, Levine WN. The rising incidence of acromioplasty. J Bone Joint Surg Am 2010;92:1842-50.
48. Ketola S, Lehtinen JT, Arnala I. Arthroscopic decompression not recommended in the treatment of rotator cuff tendinopathy: a final review of a randomised controlled trial at a minimum follow-up of ten years. Bone Jt J 2017;99-B:799-805.
49. Chester R, Khondoker M, Shepstone L, Lewis JS, Jerosch-Herold C. Self-efficacy and risk of persistent shoulder pain: results of a Classification and Regression Tree (CART) analysis. Br J Sports Med 2019;53(13):825-34.
50. Dunn WR, Kuhn JE, Sanders R. MOON Shoulder Group. 2013 Neer Award: predictors of failure of nonoperative treatment of chronic, symptomatic, full-thickness rotator cuff tears. J Shoulder Elbow Surg 2016;25(8):1303-11.
51. Dunn WR, Kuhn JE, Sanders R. Symptoms of pain do not correlate with rotator cuff tear severity: a cross-sectional study of 393 patients with a symptomatic atraumatic full-thickness rotator cuff tear. J Bone Joint Surg Am 2014;96(10):793-800.

SUPPLEMENTS

Appendix 1. Rationale and main criticisms of clinical tests and instrumental examinations.

1. Clinical tests: the objective of special tests is to make a differential diagnosis of shoulder disorders based on painful responses to provocative maneuvers; however, evidence of efficacy supporting the quality of the psychometric properties of these procedures is lacking as well as dubious, so this led the authors of recent literature to question the quality of these properties and the diagnostic value of these maneuvers (18, 19).

2. Instrumental examination: in the case of the subacromial impingement syndrome, US, MRI and MRA, which are now increasingly used, have as their objective the direct confirmation of rotator cuff lesions, specifically the tendon of the supraspinatus even if it has been amply demonstrated in numerous studies, that there is no direct correlation between structural alterations and pain; in addition, the lesions shown during imaging can be asymptomatic (4, 13, 17).

Appendix 2. Keywords, filters and eligibility criteria used for the construction of the strings.

A)

	CLINICAL TEST STRING	INSTRUMENTAL EXAMINATION STRING
POPULATION	Shoulder impingement syndrome [Mesh], Rotator cuff impingement syndrome, Rotator cuff impingement, Subacromial impingement syndrome, Rotator cuff syndrome, Subacromial pain, Shoulder Pain [Mesh], Rotator Cuff [Mesh], Supraspinatus, Infraspinatus, Teres minor, Subscapularis.	Shoulder impingement syndrome [MesH], Shoulder pain [MesH], Rotator cuff syndrome, Rotator cuff impingement, Rotator cuff impingement syndrome, Subacromial impingement syndrome [MesH], Subacromial pain, Rotator cuff [MesH], Supraspinatus, Infraspinatus, Subscapularis, Teres minor.
EXPOSURE	Test, Clinical test, Special test, Assessment test, Physical test, Neer sign, Hawkins-Kennedy test, Yocum test, Internal Rotation Resistance Strenght Test, Painful arc, Empty can test, Jobe test, Resistance, Strenght, Physical examination [Mesh], Assessment, Clinical assessment, Clinical evaluation, Clinical examination.	Diagnostic imaging [MesH], Imaging, MRI, Magnetic Resonance Imaging [MesH], Ultrasound, Ultrasonography [MesH], Magnetic Resonance Arthrography, Radiography [MesH], X-Ray, Diagnostic procedures, Diagnosis, Instrumental diagnosis.

FILTERS

1. ((systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti] OR systematic literature review [ti] OR this systematic review [tw] OR pooling project [tw] OR (systematic review [tiab] AND review [pt]) OR meta synthesis [ti] OR meta-analy*[ti] OR integrative review [tw] OR integrative research review [tw] OR rapid review [tw] OR umbrella review [tw] OR consensus development conference [pt] OR practice guideline [pt] OR drug class reviews [ti] OR cochrane database syst rev [ta] OR acp journal club [ta] OR health technol assess [ta] OR evid rep technol assess summ [ta] OR jbi database system rev implement rep [ta]) OR (clinical guideline [tw] AND management [tw]) OR ((evidence based[ti] OR evidence-based medicine [mh] OR best practice* [ti] OR evidence synthesis [tiab]) AND (review [pt] OR diseases category[mh] OR behavior and behavior mechanisms [mh] OR therapeutics [mh] OR evaluation studies[pt] OR validation studies[pt] OR guideline [pt] OR pmcbook)) OR ((systematic [tw] OR systematically [tw] OR critical [tiab] OR (study selection [tw]) OR (predetermined [tw] OR inclusion [tw] AND criteri* [tw]) OR exclusion criteri* [tw] OR main outcome measures [tw] OR standard of care [tw] OR standards of care [tw]) AND (survey [tiab] OR surveys [tiab] OR overview* [tw] OR review [tiab] OR reviews [tiab] OR search* [tw] OR handsearch [tw] OR analysis [ti] OR critique [tiab] OR appraisal [tw] OR (reduction [tw]AND (risk [mh] OR risk [tw]) AND (death OR recurrence))) AND (literature [tiab] OR articles [tiab] OR publications [tiab] OR publication [tiab] OR bibliography [tiab] OR bibliographies [tiab] OR published [tiab] OR pooled data [tw] ORunpublished [tw] OR citation [tw] OR citations [tw] OR database [tiab] OR internet [tiab] OR textbooks [tiab] OR references [tw] OR scales [tw] OR papers [tw] OR datasets [tw] OR trials [tiab] OR meta-analy* [tw] OR (clinical [tiab] AND studies [tiab]) OR treatment outcome [mh] OR treatment outcome [tw] OR pmcbook)) NOT (letter [pt] OR newspaper article [pt]))

2. Humans

3. English

B)

ELIGIBILITY CRITERIA

CLINICAL TEST STRING

Inclusion criteria

Revisions that exclusively concern the use of clinical tests in the evaluation of subacromial shoulder impingement have been included. Only systematic reviews in English have been included.

Exclusion criteria

Studies concerning other components of the diagnostic path were excluded, such as: medical history collection, observation, diagnostic imaging, evaluation of ROM and muscle strength. In addition, studies concerning surgery or post-surgery and involving other shoulder disorders have been excluded.

INSTRUMENTAL EXAMINATIONS STRING

Inclusion criteria

Revisions concerning only the use of US, MRI and MRA in the diagnosis of subacromial impingement have been included. Only systematic reviews in English have been selected.

Exclusion criteria

Non-English articles, articles related to other shoulder disorders and post-surgical studies have been excluded.

Appendix 3. Search strings.

1. Clinical test search string: ("Shoulder impingement syndrome" [Mesh] OR "Rotator cuff impingement syndrome" OR "Rotator cuff impingement" OR "Subacromial impingement syndrome" OR "Rotator cuff syndrome" OR "Subacromial pain" OR "Shoulder Pain" [Mesh] OR "Rotator Cuff" [Mesh] OR "Supraspinatus" OR "Infraspinatus" OR "Teres minor" OR "Subscapularis") AND ("Test" OR "Clinical test" OR "Special test" OR "Assessment test" OR "Physical test" OR "Neer sign" OR "Hawkins-Kennedy test" OR "Yocum test" OR "Internal Rotation Resistance Strenght Test" OR "Painful arc" OR "Empty can test" OR "Jobe test" OR "Resistance" OR "Strenght" OR "Physical examination" [Mesh] OR "Assessment" OR "Clinical assessment" OR "Clinical evaluation" OR "Clinical examination") AND ((systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti] OR systematic literature review [ti] OR this systematic review [tw] OR pooling project [tw] OR (systematic review [tiab] AND review [pt]) OR meta synthesis [ti] OR meta-analy*[ti] OR integrative review [tw] OR integrative research review [tw] OR rapid review [tw] OR umbrella review [tw] OR consensus development conference [pt] OR practice guideline [pt] OR drug class reviews [ti] OR cochrane database syst rev [ta] OR acp journal club [ta] OR health technol assess [ta] OR evid rep technol assess summ [ta] OR jbi database system rev implement rep [ta]) OR (clinical guideline [tw] AND management [tw]) OR ((evidence based[ti] OR evidence-based medicine [mh] OR best practice* [ti] OR evidence synthesis [tiab]) AND (review [pt] OR diseases category[mh] OR behavior and behavior mechanisms [mh] OR therapeutics [mh] OR evaluation studies[pt] OR validation studies[pt] OR guideline [pt] OR pmcbook)) OR ((systematic [tw] OR systematically [tw] OR critical [tiab] OR (study selection [tw]) OR (predetermined [tw] OR inclusion [tw] AND criteri* [tw]) OR exclusion criteri* [tw] OR main outcome measures [tw] OR standard of care [tw] OR standards of care [tw]) AND (survey [tiab] OR surveys [tiab] OR overview* [tw] OR review [tiab] OR reviews [tiab] OR search* [tw] OR handsearch [tw] OR analysis [ti] OR critique [tiab] OR appraisal [tw] OR (reduction [tw]AND (risk [mh] OR risk [tw]) AND (death OR recurrence))) AND (literature [tiab] OR articles [tiab] OR publications [tiab] OR publication [tiab] OR bibliography [tiab] OR bibliographies [tiab] OR published [tiab] OR pooled data [tw] ORunpublished [tw] OR citation [tw] OR citations [tw] OR database [tiab] OR internet [tiab] OR textbooks [tiab]

OR references [tw] OR scales [tw] OR papers [tw] OR datasets [tw] OR trials [tiab] OR meta-analy* [tw] OR (clinical [tiab] AND studies [tiab]) OR treatment outcome [mh] OR treatment outcome [tw] OR pmcbook)) NOT (letter [pt] OR newspaper article [pt])).

2) Instrumental examination search string: ((“shoulder pain”[MesH Terms] OR “rotator cuff syndrome”[All Fields] OR “rotator cuff impingement”[All Fields] OR “rotator cuff impingement syndrome”[All Fields] OR “subacromial pain”[All Fields] OR “rotator cuff”[MesH Terms] OR “supraspinatus”[All Fields] OR “infraspinatus”[All Fields] OR “subscapularis”[All Fields] OR “teres minor”[All Fields] AND (“diagnostic imaging”[MeSH Terms] OR “imaging”[All Fields] OR “MRI”[All Fields] OR “magnetic resonance imaging”[MesH Terms] OR “ultrasound”[All Fields] OR “ultrasonography”[MesH Terms] OR “magnetic resonance arthrography”[All Fields] OR “diagnostic procedures”[All Fields] OR “radiography”[MesH Terms] OR “x-ray”[All Fields] OR “diagnosis”[All Fields] OR “instrumental diagnosis”[All Fields]))) AND ((systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti] OR systematic literature review [ti] OR this systematic review [tw] OR pooling project [tw] OR (systematic review [tiab] AND review [pt]) OR meta synthesis [ti] OR meta-analy*[ti] OR integrative review [tw] OR integrative research review [tw] OR rapid review [tw] OR umbrella review [tw] OR consensus development conference [pt] OR practice guideline [pt] OR drug class reviews [ti] OR cochrane database syst rev [ta] OR acp journal club [ta] OR health technol assess [ta] OR evid rep technol assess summ [ta] OR jbi database system rev implement rep [ta]) OR (clinical guideline [tw] AND management [tw]) OR ((evidence based[ti] OR evidence-based medicine [mh] OR best practice* [ti] OR evidence synthesis [tiab]) AND (review [pt] OR diseases category[mh] OR behavior and behavior mechanisms [mh] OR therapeutics [mh] OR evaluation studies[pt] OR validation studies[pt] OR guideline [pt] OR pmcbook)) OR ((systematic [tw] OR systematically [tw] OR critical [tiab] OR (study selection [tw]) OR (predetermined [tw] OR inclusion [tw] AND criteri* [tw]) OR exclusion criteri* [tw] OR main outcome measures [tw] OR standard of care [tw] OR standards of care [tw]) AND (survey [tiab] OR surveys [tiab] OR overview* [tw] OR review [tiab] OR reviews [tiab] OR search* [tw] OR handsearch [tw] OR analysis [ti] OR critique [tiab] OR appraisal [tw] OR (reduction [tw]AND (risk [mh] OR risk [tw]) AND (death OR recurrence))) AND (literature [tiab] OR articles [tiab] OR publications [tiab] OR publication [tiab] OR bibliography [tiab] OR bibliographies [tiab] OR published [tiab] OR pooled data [tw] OR unpublished [tw] OR citation [tw] OR citations [tw] OR database [tiab] OR internet [tiab] OR textbooks [tiab] OR references [tw] OR scales [tw] OR papers [tw] OR datasets [tw] OR trials [tiab] OR meta-analy* [tw] OR (clinical [tiab] AND studies [tiab]) OR treatment outcome [mh] OR treatment outcome [tw] OR pmcbook)) NOT (letter [pt] OR newspaper article [pt]))

Appendix 4. AMSTAR 2.

RATING OVERALL CONFIDENCE IN THE RESULTS OF THE REVIEW
HIGH QUALITY No or one non-critical weakness: the systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest.
MODERATE QUALITY More than one non-critical weakness*: the systematic review has more than one weakness but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review.
LOW QUALITY One critical flaw with or without non-critical weaknesses: the review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest.
CRITICALLY LOW QUALITY More than one critical flaw with or without non-critical weaknesses: the review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies.
*Multiple non-critical weaknesses may diminish confidence in the review and it may be appropriate to move the overall appraisal down from moderate to low confidence.

Appendix 5. AMSTAR 2 score.

A)
CLINICAL TESTS
Legend: Y: YES; PY: Partial YES; N: NO; RCTs: Randomised controlled trial study; NRSI: Non-randomised studies of interventions; NM: A meta-analysis was not conducted.

ARTICLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	SCORE
1. Gismervik *et al.*, 2017	Y	Y	N	PY	Y	Y	Y	Y	RCTs Y	N	Y	Y	Y	Y	N	Y	Low Quality
2. Lange *et al.*, 2016	Y	PY	N	Y	Y	Y	N	N	N	N	Y	Y	N	Y	N	N	Critically Low Quality
3. O'Kane *et al.*, 2014	Y	PY	N	PY	N	N	N	PY	RCTs Y	N	N	Y	Y	N	N	Y	Critically Low Quality
4. Hermans *et al.*, 2013	Y	PY	Y	PY	Y	Y	N	PY	RCTs Y	N	Y	N	N	N	N	Y	Critically Low Quality
5. Hanchard *et al.*, 2013	Y	PY	Y	PY	Y	Y	Y	Y	Y	N	NM	NM	N	N	NM	N	Low Quality
6. Hegedus *et al.*, 2008	Y	Y	N	PY	Y	Y	N	N	Y	N	Y	N	N	Y	N	Y	Low Quality
7. Hegedus *et al.*, 2012	Y	Y	N	PY	Y	Y	N	PY	Y	N	Y	Y	Y	Y	Y	Y	Moderate Quality
8. Alqunaee *et al.*, 2012	Y	Y	N	PY	Y	Y	N	PY	NRSI Y	N	Y	Y	Y	N	N	Y	Low Quality
9. May *et al.*, 2010	Y	PY	N	PY	Y	Y	N	N	Y	N	NM	NM	N	N	N	Y	Low Quality
10. Beaudreuil *et al.*, 2009	Y	N	N	PY	N	N	N	N	N	N	NM	NM	N	N	N	N	Critically Low Quality
11. Hughes *et al.*, 2008	Y	PY	N	PY	Y	Y	Y	N	N	N	NM	NM	Y	N	N	N	Critically Low Quality
12. Dinnes *et al.*, 2003	Y	Y	Y	PY	Y	Y	Y	Y	NRSI Y	N	NRSI Y	N	Y	Y	Y	Y	Low Quality

Legend: Y: YES; PY: Partial YES; N: NO; RCTs: Randomised controlled trial study; NRSI: Non-randomised studies of interventions; NM: A meta-analysis was not conducted.

B)
INSTRUMENTAL EXAMINATIONS
Y: YES; PY: Partial YES; N: NO; RCTs: Randomised controlled trial study; NRSI: Non-randomised studies of interventions; NM: A meta-analysis was not conducted.

ARTICLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	SCORE
1. Mc Garvey *et al.*, 2016	Y	PY	N	PY	N	N	N	N	NRSI N	N	NN	Y	N	Y	N	Y	Critically Low Quality
2. Roy *et al.*, 2015	Y	N	N	N	N	N	N	PY	NRSI N	N	NY	Y	N	N	N	Y	Critically Low Quality
3. Lenza *et al.*, 2013	Y	Y	N	PY	Y	Y	Y	Y	NRSI Y	N	NY	Y	N	Y	N	Y	Critically Low Quality
4. Amee L. Seitz *et al.*, 2010	Y	PY	Y	PY	N	Y	N	PY	NRSI PY	N	NY	Y	Y	Y	N	N	Low Quality
5. Ottenheijm *et al.*, 2010	Y	N	N	PY	Y	Y	N	PY	NRSI N	N	NY	Y	N	Y	N	N	Critically Low Quality
6. De Jesus *et al.*, 2009	Y	N	N	N	N	N	N	N	NRSI N	N	NY	Y	N	Y	Y	N	Critically Low Quality
7. Dinnes *et al.*, 2003	Y	Y	N	PY	Y	Y	Y	PY	NRSI Y	N	NY	Y	Y	Y	Y	Y	Low Quality

Y: YES; PY: Partial YES; N: NO; RCTs: Randomised controlled trial study; NRSI: Non-randomised studies of interventions; NM: A meta-analysis was not conducted.