Health Technology Assessment 2003; Vol. 7: No. 29

The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review

J Dinnes
E Loveman
L McIntyre
N Waugh

Health Technology Assessment
NHS R&D HTA Programme

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review

J Dinnes*
E Loveman
L McIntyre
N Waugh

Southampton Health Technology Assessments Centre,
University of Southampton, UK

*Corresponding author

Declared competing interests of authors: none

Published October 2003

This report should be referenced as follows:

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).

Health Technology Assessment is indexed in *Index Medicus*/MEDLINE and *Excerpta Medica*/EMBASE.

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and provide care in the NHS.

The research reported in this monograph was commissioned by the HTA Programme and was funded as project number 01/39/01. Technology assessment reports are completed in a limited time to inform decisions in key areas by bringing together evidence on the use of the technology concerned.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series
Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report.

ISSN 1366-5278

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.
Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.

Abstract

The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review

J Dinnes,* E Loveman, L McIntyre and N Waugh

Southampton Health Technology Assessments Centre, University of Southampton, UK
*Corresponding author

Objectives: To evaluate the evidence for the effectiveness and cost-effectiveness of the newer diagnostic imaging tests as an addition to clinical examination and patient history for the diagnosis of soft tissue shoulder disorders.
Data sources: Literature was identified from several sources including general medical databases.
Review methods: Studies were identified that evaluated clinical examination, ultrasound, magnetic resonance imaging (MRI), or magnetic resonance arthrography (MRA) in patients suspected of having soft tissue shoulder disorders. Outcomes assessed were clinical impingement syndrome or rotator cuff tear (full, partial or any). Only cohort studies were included. The methodological quality of included test accuracy studies was assessed using a formal quality assessment tool for diagnostic studies and the extraction of study findings was conducted in duplicate using a pre-designed and piloted data extraction form to avoid any errors. For each test, sensitivity, specificity and positive and negative likelihood ratios with 95% confidence intervals were calculated for each study. Where possible pooled estimates of sensitivity, specificity and likelihood ratios were calculated using random effects methods. Potential sources of heterogeneity were investigated by conducting subgroup analyses.
Results: In the included studies, the prevalence of rotator cuff disorders was generally high, partial verification of patients was common and in many cases patients who were selected retrospectively because they had undergone the reference test. Sample sizes were generally very small. Reference tests were often inappropriate with many studies using arthrography alone, despite problems with its sensitivity. For clinical assessment, 10 cohort studies were found that examined either the accuracy of individual tests or clinical examination as a whole: individual tests were either good at ruling out rotator cuff tears when negative (high sensitivity) or at ruling in such disorders when positive (high specificity), but small sample sizes meant that there was no conclusive evidence. Ultrasound was investigated in 38 cohort studies and found to be most accurate when used for the detection of full-thickness tears; sensitivity was lower for detection of partial-thickness tears. For MRI, 29 cohort studies were included. For full-thickness tears, overall pooled sensitivities and specificities were fairly high and the studies were not statistically heterogeneous; however for the detection of partial-thickness rotator cuff tears, the pooled sensitivity estimate was much lower. The results from six MRA studies suggested that it may be very accurate for detection of full-thickness rotator cuff tears, although its performance for the detection of partial-thickness tears was less consistent. Direct evidence for the performance of one test compared with another is very limited.
Conclusions: The results suggest that clinical examination by specialists can rule out the presence of a rotator cuff tear, and that either MRI or ultrasound could equally be used for detection of full-thickness rotator cuff tears, although ultrasound may be better at picking up partial tears. Ultrasound also may be more cost-effective in a specialist hospital setting for identification of full-thickness tears. Further research suggestions include the need for large, well-designed, prospective studies of the diagnosis of shoulder pain, in particular a follow-up study of patients with shoulder pain in primary care and a prospective cohort study of clinical examination, ultrasound and MRI, alone and/or in combination.

Contents

Glossary and list of abbreviations

Glossary

Adhesive capsulitis Occurs where inflammation and pain limit the use of the shoulder to the extent that the joint stiffens.

Arthrography Examination of a joint by X-ray after injection of a contrast medium.

Arthroscopy An invasive endoscopic examination of a joint.

Bursitis Occurs where repeated use of the shoulder leads to inflammation and swelling of a bursa.

Clinical impingement syndrome A spectrum of shoulder pathologies with distinct clinical symptoms such as pain on attempted use of the shoulder.

Magnetic resonance arthrography Examination of a joint by MRI after injection of a contrast medium.

Magnetic resonance imaging A non-invasive method of demonstrating internal anatomy based on the use of magnetic fields which are translated into a computerised image.

Referred pain Pain from deep structures perceived as arising from a surface area remote from its actual origin.

Tendinitis Inflammation of tendons and of tendon muscle attachments.

Tendinosis Degeneration of tendons and of tendon muscle attachments.

Ultrasound A non-invasive imaging technique which visualises deep structures of the body by recording the reflections of echoes of pulses of ultrasonic waves directed into the tissues.

List of abbreviations

AC	angled coronal
AS	angled sagittal
AX	axial
CCS	clinical community setting
CI	confidence interval
CTA	computed tomographic arthrography
df	degrees of freedom
2D	two-dimensional
3D	three-dimensional
ERLS	external rotation lag sign
FN	false negative
FoV	field of view
FSE	fast spin-echo
GRE	gradient echo
IRLS	internal rotation lag sign
IRRST	internal rotation resistance strength test
LR	likelihood ratio
MPGR	multiplanar gradient-recalled
MR	magnetic resonance
MRA	magnetic resonance arthrography
MRI	magnetic resonance imaging
NSAID	non-steroidal anti-inflammatory drug
OC	oblique coronal

continued

List of abbreviations *continued*

OP	oblique parasagittal
OS	oblique sagittal
PC	paracoronal
PS	parasagittal
RC	rotator cuff
RCT	rotator cuff tear
SA	subacromial
Sag	sagittal
SD	subdeltoid
SIS	subacromial impingement syndrome
SIT	subacromial injection test
STIR	short TI inversion recovery
TE	echo time
Ten	tendinitis
TI	inversion time
TP	true positive
TR	repetition time

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Background

Shoulder pain is a significant cause of morbidity; the prevalence of self-reported pain is estimated to be between 16 and 26%, and it is the third most common cause of musculoskeletal consultation in primary care. The cause can be difficult to diagnose owing to the complex anatomy of the shoulder and the spectrum of underlying disorders. Most shoulder problems fall into three major categories: soft tissue disorders, articular injury or instability, and arthritis. The incidence of lesions increases with age as tendon tissue progressively weakens or degenerates, but repeated microtrauma or overuse from professional or athletic activity can also cause soft tissue problems in all age groups.

There are no clear national guidelines for the diagnosis of shoulder pain. Several diagnostic tests are used for the diagnosis of soft tissue disorders, including clinical assessment, ultrasonography, magnetic resonance imaging (MRI), magnetic resonance arthrography (MRA) and arthroscopy, yet their relative accuracy, cost-effectiveness and impact on quality of life are uncertain.

Objective

To evaluate the evidence for the effectiveness and cost-effectiveness of the newer diagnostic imaging tests as an addition to clinical examination and patient history for the diagnosis of soft tissue shoulder disorders.

Methods

Literature search

Literature was identified from several sources including general medical databases.

Inclusion criteria

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.

Quality assessment and data extraction

The methodological quality of included test accuracy studies was assessed using a formal quality assessment tool for diagnostic studies developed by the NHS Centre for Reviews and Dissemination at the University of York. The extraction of study findings was conducted in duplicate using a pre-designed and piloted data extraction form to avoid any errors.

Data synthesis

For each test, sensitivity, specificity and positive and negative likelihood ratios (LRs) with 95% confidence intervals were calculated for each study. Where no trade-off between sensitivity and specificity was revealed, and studies were otherwise sufficiently homogeneous, pooled estimates of sensitivity, specificity and LRs were calculated using random effects methods. Potential sources of heterogeneity were investigated by conducting subgroup analyses according to features of the population (spectrum), test, and reference test, and study quality.

Results

Quality of included studies

The prevalence of rotator cuff (RC) disorders was high in most studies (overall mean prevalence over 50% for all tests), although it varied according to the setting and outcome used. The study setting was not always reported, but where it was, only two were conducted in centres other than hospital radiology or orthopaedics departments. Partial verification of patients was common and in many studies patients were selected because they had undergone the reference test. Sample sizes were generally very small, with overall means of less than 100.

The reference tests used in the studies were often inappropriate with many studies (especially ultrasound studies) using arthrography alone, despite problems with its sensitivity. Others used more than one reference test, in some cases clearly stating that the test used was based on the result of the index test.

Few studies reported details of those interpreting the tests other than that they were orthopaedists or radiologists, often specialising in shoulder disorders.

Clinical assessment

Ten cohort studies were included: seven examined the accuracy of individual clinical examination tests and six estimated the accuracy of clinical examination *per se* or the combination of two or more positive test results. Individual tests were either good at ruling out RCTs when negative (high sensitivity) or at ruling in such disorders when positive (high specificity), but small sample sizes mean that there was no conclusive evidence for any single test that can conclusively diagnose RC disorders. Pooled results from four studies that evaluated clinical examination as a whole indicated overall sensitivity and specificity to be 0.90 (95% CI: 0.87 to 0.93) and 0.54 (95% CI: 0.47 to 0.61) for detection of full-thickness RCTs.

Ultrasound

Thirty-eight cohort studies investigating the accuracy of ultrasound were identified. Ultrasound was most accurate when used for the detection of full-thickness tears, although results were heterogeneous: pooled sensitivity 0.87 (95% CI: 0.84 to 0.89) and specificity 0.96 (0.94 to 0.97). Sensitivity was lower for detection of partial-thickness tears (0.67, 95% CI: 0.61 to 0.73) although specificity remained high, and studies were again very heterogeneous. Statistically, several possible reasons for the differences in sensitivity estimates between studies were identified, including prevalence and mean age. The number of studies available limited the power of the subgroup analyses. It remains to be determined whether or not ultrasound can provide such conclusive evidence for the value of a negative ultrasound finding in ruling out the presence of a tear.

MRI

Twenty-nine cohort studies were included, most using conventional MRI pulse sequences as opposed to fat-suppressed MRI. For full-thickness tears, overall pooled sensitivities and specificities were high (0.89, 95% CI: 0.86 to 0.92; and 0.93, 95% CI: 0.91 to 0.95, respectively) and the studies were not statistically heterogeneous. For detection of partial-thickness RCTs, pooled sensitivity estimate was much lower (0.44, 95% CI: 0.36 to 0.51) although specificity again remained high (0.90, 95% CI: 0.87 to 0.92). Where tear prevalence is relatively high, a negative magnetic resonance finding may be sufficient to rule out the presence of a full-thickness tear, but between study heterogeneity means that similar conclusions cannot yet be drawn regarding a positive test result.

MRA

Six studies investigating the accuracy of MRA were included. The type of MRI, views and contrast used varied considerably between studies, making any conclusions difficult. The pooled results suggest that MRA may be very accurate for detection of full-thickness RCTs [overall pooled sensitivity 0.95 (95% CI: 0.82 to 0.98) and specificity 0.93 (95% CI: 0.84 to 0.97), both estimates homogeneous]. Its performance for the detection of partial-thickness tears is less consistent. There is also some suggestion that MRA performs better than ultrasound or MRI, but any such benefit must be set against the invasiveness and potential discomfort to patients of the procedure

Direct comparisons of tests

Direct evidence for the performance of one test compared with another is very limited. Further research is needed to determine the place of these imaging tests in the diagnosis of RC disorders.

Conclusions

Our results suggest that clinical examination by specialists can rule out the presence of a RCT, and that either MRI **or** ultrasound could equally be used for detection of full-thickness RCTs. Although still not by any means accurate, ultrasound **may** be better at picking up partial tears. Given the large differential in the cost of the two procedures, the implication from current evidence is that ultrasound is the more cost-effective test to use in a specialist hospital setting for identification of full-thickness tears. Whether or not these results are transferable to settings with lower prevalence, different spectra of disease and less-specialised clinicians, such as in primary care, remains to be determined.

Implications for further research

There is a need for large, well-designed, prospective studies of the diagnosis of shoulder pain. In particular, a follow-up study of patients with shoulder pain in primary care is needed to inform our understanding of the natural history and epidemiology of shoulder pain and, for those patients referred to secondary care, a prospective cohort study of clinical examination, ultrasound and MRI, alone and/or in combination is also needed. The ability of these tests not only to diagnose the spectrum of soft tissue shoulder disorders (not just RCT) but also to inform treatment decisions remains to be determined.

Chapter 1

Aim of the review

The aim of this systematic review is to evaluate currently available evidence for the effectiveness and cost-effectiveness of the newer diagnostic imaging tests as an addition to clinical examination and patient history for the diagnosis of soft tissue shoulder disorders.

Shoulder pain is a significant cause of morbidity, and is the third most common cause of musculoskeletal consultation in primary care. The cause can be difficult to diagnose owing to the spectrum of underlying disorders, both of the shoulder joint itself and surrounding tissue, and pain due to problems elsewhere such as in the cervical spine. There are no clear national guidelines for diagnosis within secondary care. Clinical assessment, ultrasonography, magnetic resonance imaging (MRI), magnetic resonance arthrography (MRA) and arthroscopy are all used for the diagnosis of soft tissue disorders, yet their relative accuracy, cost-effectiveness and impact on quality of life are uncertain. This review was therefore commissioned in order to inform the development of any related primary HTA research.

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.

Chapter 2
Background

Shoulder pain

Epidemiology

Shoulder pain is a significant cause of morbidity in the general population. Recent surveys have estimated the prevalence of self-reported shoulder pain at 16%[1] in the UK, rising to 26%[2] in the elderly and at 21% in The Netherlands.[3] About 1% of adults over age 45 years in the UK present each year with a new episode of shoulder pain. Prevalence estimates for shoulder pain range from 4 to 20%.[4] It has been suggested that only 40–50% of people with shoulder pain will consult a general practitioner (GP) for it;[3] nevertheless it is the third most common cause of musculoskeletal consultation in primary care. Around a quarter of patients presenting with a new episode of shoulder pain report a previous episode of shoulder pain.[5]

The anatomy of the shoulder

The shoulder is made up of three bones:

- the scapula or shoulder blade – a wide, thin triangular bone with two extensions (called 'processes') at the shoulder joint, the coracoid and acromion
- the clavicle or collar-bone
- the humerus (in the upper arm)

and two joints:

- the acromio-clavicular joint (clavicle to scapula)
- the humero-scapular joint (the main shoulder joint).

For the latter joint, the ball of the humerus fits on to the glenoid 'cavity' of the scapula. Although this is classed as a ball-and-socket joint, the socket is more like a flat saucer than a cup (as in the hip joint), and the integrity of the shoulder joint therefore depends on ligaments and muscles, rather than a socket.

The joint capsule is a loose, fibrous cylinder of horizontal fibres going from scapula to humerus. It has to be loose to allow the range of movement at the shoulder. Hence the joint capsule itself also does little for stability. The anterior part of the joint capsule is thickened and strengthened by the three glenohumeral ligaments. The superoposterior part is strengthened near its humeral attachment by the coracohumeral ligament.

Of the many muscles around the shoulder, there are those of the shoulder proper (defined as those which produce shoulder movements, thereby excluding those such as biceps which are in the shoulder but which act on other joints such as the elbow), and also the adjacent muscles of upper back and upper chest (the pectoralis major and minor and the latissimus dorsi). Simplistically, the muscles initiating shoulder movement include:

- deltoid – the most powerful abductor, but only acts after the supraspinatus has brought about the first 15° or so of movement
- supraspinatus – which initiates abduction
- infraspinatus – which is responsible for lateral rotation
- teres major – adduction and medial rotation
- teres minor – lateral rotation and some adduction
- subscapularis – similar to teres major.

The combined action of the subscapularis, infraspinatus and teres minor counteracts the upward component of the pull of the deltoid, thus enabling the deltoid (aided by the supraspinatus) to create a force component perpendicular to the humeral shaft to abduct the arm.

Tendons are parts of muscles, where the fleshy part of the muscle is replaced by a ligament-like structure before it is attached to bones. The action of the intact muscle tendon units comprising the rotator cuff (RC) is very important in maintaining the stability of the glenohumeral joint while movements take place. The RC tendons include the tendons of the teres minor, infraspinatus, supraspinatus and subscapularis muscles. These tendons are directly related to the shoulder joint capsule and spread out over it, blending with it near their humeral attachments, in effect acting as extensible ligaments.

The structure of the joint therefore allows considerable movement, but at the expense of stability. Movements include:

- flexion – bringing the arm forwards (as in shaking hands)
- extension backwards (as in reaching for a hip pocket)
- abduction out sideways, away from the body
- adduction, bringing the arm back down to the side of the chest
- rotation – with the elbow held at the side of the chest, internal rotation of the humerus allows the hand to touch the opposite side of the chest.

The RC muscles contract, for example when a weight is being carried in the hand, thereby pulling the head of the humerus against the glenoid fossa and in effect protecting the shoulder from dislocating downwards. During shoulder movements, the muscles act in unison to stabilise the joint.

Causes of shoulder pain

The complex anatomy of the shoulder and the wide spectrum of disorders that can lead to shoulder pain symptoms can make accurate diagnosis difficult. The difficulty in locating the precise anatomical cause of shoulder pain is in part due to the frequent radiation of pain from neck structures into the shoulder and arm. The lack of consensus on the appropriate diagnostic criteria and the availability of several diagnostic classifications illustrate the complexity of diagnosis.[6]

Most shoulder problems fall into three major categories: soft tissue disorders, articular injury or instability, and arthritis. It has been estimated that up to 90% of lesions causing painful shoulder result from extracapsular soft tissue lesions, and these will be the main focus of this report.

Soft tissue shoulder disorder: RC disorders and impingement syndrome

Soft tissue shoulder disorders either result from the wearing process that takes place over a period of years or may occur from acute strain or repetitive use, causing friction and irritation. The term 'clinical impingement syndrome' covers a spectrum of shoulder pathologies with distinct clinical symptoms such as pain on attempted use of the shoulder, particularly with overhead activity and/or weakness. The most common of these are shoulder bursitis, RC tendinosis and rotator cuff tears (RCTs).

Bursitis occurs where repeated use of the shoulder leads to inflammation and swelling of a bursa (fluid-filled sacs located around the joints that lessen the friction caused by movement of the shoulder). Bursitis often occurs in association with rotator cuff tendinitis and tends to be triggered by sports activities, or more commonly physical activity.[7]

RC disorders most commonly occur where tendon tissue has weakened or degenerated due to ageing, hypovascularity of the critical zone, repeated microtrauma or overuse from professional or athletic activity.[8] RC tendinitis is more often seen in middle-aged or older patients with chronic shoulder pain than in younger people. Given the key role of the RC muscles in stabilising the joint, any lesion is likely to cause pain. The best known syndrome is supraspinatus tendinitis with its classical arc. This occurs because during abduction of the shoulder, the supraspinatus tendon would rub against the acromion were it not for the intervening and lubricating sub-acromial bursa. If this bursa degenerates, friction results. Pain can also result if the tendon is swollen due to inflammation, since the space is then too small.

RCT is the end stage on the impingement syndrome spectrum and is more painful than the most severe form of RC tendinitis. RCTs may be either partial or full-thickness tears. This splitting and tearing of tendons may result from an injury or degenerative changes in the tendons due to advancing age. Full-thickness tears also tend to be more common in older people. Anteriorly located lesions (subscapularis) are more likely associated with trauma than posteriorly located ones (infraspinatus).[9] There are three main types of partial-thickness tears according to location:

1. inferior tears at the articular surface
2. superior tears at the bursal surface
3. tears within the tendon substance.

Partial tears are more difficult to diagnose, especially very small ones (which can be mistaken for tendinopathy) and large ones (which can be mistaken for full tears).

Unfortunately, a lack of consensus regarding diagnostic criteria and the classification of shoulder disorders makes it difficult to estimate the frequency of the underlying causes of shoulder pain. A systematic review of the treatment of shoulder pain found that 24 studies had specified 10 distinct diagnoses with 16 definitions to characterise their study population.[10] In one primary study, up to 30% of soft tissue disorders were attributed to RC tendinitis, 22% to capsular

syndrome, 17% to acute bursitis and 13% to chronic bursitis.[11] Another found that around 70% of shoulder disorders primarily involved the RC.[2] The difficulties of diagnosis are further emphasised by evidence that at the primary care level at least, initial diagnostic categories can change significantly with follow-up visits over time.[11,12]

Other causes of shoulder pain

Injury or instability

Dislocation can occur when the head of the humerus is displaced from the glenoid cavity through injury or instability and can be combined with fracture to give fracture dislocation. Fractures and damage to articular cartilage can also occur in association with dislocation. Once the dislocation is reduced, the patient usually recovers in a couple of weeks.[7] Instability can result in a tendency to repeated dislocation of one of the shoulder joints. Recurrent instability is closely correlated with age: patients less than 25 years old at the time of the initial episode have a recurrence rate of 60–90% compared with only 15% in those aged over 45 years.[7] Although the dislocations are easy to manipulate back into position, usually under general anaesthetic, the shoulder joint may be left permanently weakened. This weakness is partly caused by stretching and tearing of the surrounding ligaments, tendons and capsule of the joint. There may also be damage to the cartilage and bone lining the rim of the socket.

Adhesive capsulitis

Adhesive capsulitis or 'frozen shoulder' can result when inflammation and pain limit the use of the shoulder to the extent that the joint stiffens. It is most commonly seen in older patients whose shoulders are immobilised after an injury and can occur even after minimal trauma.

Arthritis

Shoulder pain can also result from osteoarthritis or degenerative arthritis. Osteoarthritis (degenerative arthritis) involves wear and tear changes with inflammation of the joint, causing swelling, pain and stiffness. Degenerative arthritis may be related to sports or work injuries or can occur as part of the gradual ageing process. Rheumatoid arthritis is associated with a severe inflammation of the joint lining (the synovium), which then leads to breakdown of the joint surface cartilage and bone.

Referred pain

Shoulder pain can also be caused by referred pain, or radiation of pain, from internal organs, neurological or vascular disorders, neoplasms and disorders of the cervical spine.[11]

Diagnosis and treatment of soft tissue shoulder disorders

In the absence of trauma or arthritis, most shoulder pain is caused by periarticular soft tissue injury as described above. Differential diagnosis can be difficult and the most important criterion for the assessment of different imaging modalities is their ability to distinguish individual pathologies of the shoulder joint, either alone or in combination.[13] Determining the source of the problem in the shoulder is important to be able to recommend the right treatment. Most patients presenting with shoulder pain will be diagnosed and treated within general practice: only around 10% of patients seen in general practice are referred for a specialist opinion.[11]

There appear to be no UK national guidelines for the diagnosis of shoulder pain. Available tests include clinical assessment, plain X-ray, arthroscopy, ultrasonography, MRI and MRA [computed tomographic arthrography (CTA) is also used for the investigation of shoulder pain, but is not useful for soft tissue shoulder disorders].

Clinical assessment and patient history

Physical examination and history taking are the cornerstones of the diagnosis of shoulder disorders. Patient history provides the first clues to the source of the problem, can distinguish whether the problem is acute or chronic and should identify cases of referred pain from other sites such as the cervical spine.[14]

Clinical examination includes inspection and palpation, assessment of range of motion and strength and provocative shoulder testing for possible impingement syndrome and glenohumeral instability.[14] A variety of tests can be used during the examination (*Table 1*). It is thought that positive findings from these tests may be suggestive of different underlying shoulder disorders, but the individual contribution of each of these tests to differential shoulder pain diagnosis and the most accurate combination or sequencing of tests is unclear. The different diagnostic classification systems in use and the variation in classification of complaints between clinicians complicate the assessment of these tests.[6]

TABLE I *Examples of tests used in clinical examination of the shoulder*

Test	Manoeuvre	Suggested diagnosis on positive finding
Apley scratch test	Patient touches superior and inferior aspects of opposite scapula	Loss of range of motion: RC problem
Neer's sign	Arm in full flexion	Subacromial impingement
Hawkins' test	Forward flexion of the shoulder to 90° and internal rotation	Supraspinatus tendon impingement
Drop-arm test	Arm lowered slowly to waist	RCT
Cross-arm test	Forward elevation to 90° and active adduction	Acromioclavicular joint arthritis
Spurling's test	Spine extended with head rotated to affected shoulder while axially loaded	Cervical nerve root disorder
Apprehension test	Anterior pressure on the humerus with external rotation	Anterior glenohumeral instability
Relocation test	Posterior force on humerus while externally rotating the arm	Anterior glenohumeral instability
Sulcus sign	Pulling downward on elbow or wrist	Inferior glenohumeral instability
Yergason test	Elbow flexed to 90° with forearm pronated	Biceps tendon instability or tendonitis
Speed's manoeuvre	Elbow flexed 20–30° and forearm supinated	Biceps tendon instability or tendonitis
'Clunk' sign	Rotation of loaded shoulder from extension to forward flexion	Labral disorder

Adapted from Woodward and Best, 2000.[14]

Ultrasonography

Ultrasonography uses a pulse echo device to record reflected waves of a sound beam in two dimensions. It is simple, rapid, non-invasive, frequently available and relatively inexpensive. Unlike plain X-rays, it may reveal soft tissue changes.

Technique (from Zanetti and Hodler, 2000[9])

Most commonly, linear transducers with frequencies in the range 5–13 MHz are used. A recommendation to use one with ⩾7.5 MHz is often made.[13] Patients are usually examined in the upright (seated) position so that the arm can be externally rotated and placed behind the patient's back. Standardised examination protocols include examination of the acromioclavicular joint and transverse and longitudinal scans of the long biceps tendon, subscapularis, supraspinatus and infraspinatus tendons. Comparison with the opposite side is recommended. Dynamic examination (with patient moving his/her arm) is often conducted. The subscapularis can be differentiated sonographically from the supraspinatus tendon because they are separated by the long biceps tendon and also based on their course. The supraspinatus tendon, however, cannot be easily differentiated from the infraspinatus (even during arthroscopy). Ultrasound literature generally focuses on lesions of the supraspinatus tendon but rarely on the subscapularis. Diagnostic difficulties with the latter may relate to the fact that tendon degeneration and tears are commonly limited to the cranial border of the tendon; complete detachment is rarer. Resulting sonographic signs may be subtle and easily missed.

Criteria for interpretation of ultrasound scans

Normal tendon

Normal RC tendons appear echogenic (produces an acoustic shadow) compared with the adjacent deltoid muscle. The normal RC is slightly hyperechoic (produces stronger echoes) compared with the overlying muscle. The RC assumes a convex curvilinear course when it passes over the humeral head.

Mistakes can be made if the transducer is not placed directly perpendicular to the tendon – even minor deviations result in a relevant loss of echogenicity, which may be mistaken for a tendon tear.

Full-thickness RCTs
Sonographic findings include:

- non-visualisation of cuff tissue
- localised hypoechoic zones throughout the entire cuff thickness
- loss of convexity from the outer contour.

Tears may become more conspicuous when the transducer compresses the deltoid muscle against the humeral head.

Partial-thickness RCTs
Sonography findings include:

- hypoechoic defect that involves the articular or bursal surface but not the entire cuff thickness
- thinning of the cuff
- straight outer cuff border with loss of convexity.

The last two overlap with those for full-thickness tears. Other criteria, such as presence of echogenic foci, are not reliable and may be caused by tendon degeneration or calcification.

MRI

MRI is a non-invasive method of imaging, where the body is placed in a magnetic field which causes certain atomic nuclei to align in the direction of the field. Pulses of radiofrequency radiation are then applied. Interpretation of the frequencies absorbed and re-emitted allows an image in any body plane to be built up. MRI is unique in that it allows multiplanar imaging and may be of particular benefit in the differentiation of soft tissue structures.[15] As a result, MRI has been intensely investigated for the assessment of shoulder diseases including RC disease and shoulder joint instability.

MRI technique (from Lee and Lang, 2000[15])

Patients are positioned supine, with arms at their sides, in neutral position. A dedicated shoulder coil is positioned so that the lesser tuberosity of the humerus is in the centre of the coil. Three standard imaging planes are used:

1. Scout images are taken in the **axial** plane.
2. The **oblique coronal** plane is selected so that it is parallel to the course of the supraspinatus muscle. It best depicts the supraspinatus tendon, the subacromial bursa, the undersurface of the acromion, the acromioclavicular joint and the superior and inferior portions of the glenoid labrum
3. The **oblique sagittal** plane is located perpendicular to the supraspinatus muscle. These images best assess the shape and slope of the lateral acromion and its relationship to the supraspinatus tendon and muscle and the presence of fluid in the subacromial–subdeltoid bursae.

A slice thickness of 4 mm is typical, with a 1-mm interslice gap, an imaging matrix of 256 × 192 and a 12–16-cm field of view.

Possible pulse sequence protocols include:

- T1-weighted spin-echo imaging
- T2-weighted fast spin-echo (FSE) imaging with or without fat saturation.

T1-weighted images have a shorter repetition and echo time than T2-weighted images and give better definition of tendon and bone, whereas T2-weighted images are thought to easily identify the presence of fluid in the area of a tear (areas of very high signal intensity). Intermediate (T1-weighted) images can be used to delineate areas of questionable pathology and corresponding T2-weighted images can be used to identify any high signal intensity areas within the area of abnormal tendon to identify RCTs. Fat saturation is said to further increase the visibility of fluid on T2-weighted images, to reduce the signal from adipose tissue that might confound image interpretation and to reduce artefacts related to respiratory motion. Disadvantages include a decreased signal-to-noise ratio and uneven fat suppression resulting from field inhomogeneity.

FSE and gradient spin-echo pulse sequences allow faster imaging times. Gradient echo images in particular allow a significant reduction in image acquisition time by using gradient refocusing and reduced flip angles. This may provide higher contrast between the RC tendons and adjacent soft tissue than do conventional spin-echo proton density and T2-weighted sequences. Disadvantages include increased magnetic susceptibility compared with spin-echo techniques because of the lack of a 180° refocusing pulse and the potential for 'magic angle' intermediate signal intensity within the supraspinatus tendon because of the short time to echo of the technique.

Interpretation of MRI

The normal RC generally exhibits low signal intensity on MRI. The 'magic angle' phenomenon may lead to a mild increased signal in the critical zone on T1-weighted images (or proton density weighted images), but the morphology of the tendon remains normal.

Among the first authors to produce diagnostic criteria for the detection of RCTs on magnetic resonance (MR) were Zlatkin and colleagues[16] and Iannotti and colleagues.[17] Diagnosis was based on the appearance of the RC tendons (grading system) and the presence or absence of signs denoting involvement of the subacromial bursa and subacromial–subdeltoid fat plane.

RC tendons are first assigned to one of four grades:

- grade 0 – normal in signal intensity and morphology
- grade 1 – increased signal intensity with normal morphology
- grade 2 – both abnormal signal intensity and morphology; abnormal morphological findings include obvious thinning or irregularity of the tendons
- grade 3 – large and definite area of discontinuity or a gap in the normal signal void; area of discontinuity is typically seen as an area of increased signal intensity on T2-weighted images.

The peribursal fat plane, an area of high signal intensity on T1-weighted images, is found distal to the acromion and the deltoid muscle and proximal to the RC tendons. Where RCT is present, this area of high signal intensity may be obliterated on T1- and proton density-weighted images. On T2-weighted images the signal intensity consistent with that of fluid may be seen in the bursa itself.[17]

Zlatkin and colleagues[16] did not differentiate full and partial tears, but suggested that a tear is present where there is a grade 2 or 3 tendon with loss of overlying high signal intensity of the subacromial–subdeltoid fat plane on T1-weighted or proton density-weighted images and/or when fluid in the subacromial bursa was seen on T2-weighted images.[16,17]

Full-thickness tears
Accurate diagnosis is primarily based on marked contour abnormalities and associated secondary findings such as intrabursal fluid and muscle retraction. The most specific signs of full-thickness tears are:

- a gap within the tendon (interruption of tendon continuity) on T1-weighted images that becomes brighter on T2-weighted images
- retraction of the musculotendinous junction.

Other signs include the presence of free fluid in the subacromial–subdeltoid bursa, fluid in the glenohumeral joint, obliteration of the peribursal fat stripe and fatty atrophy of the muscle. Although sensitive, these signs lack specificity in that fluid may be present without RCT and not all complete RCTs have fluid. The loss of peribursal fat stripe in RCT is related to bursal fluid, granulation tissue or scar formation, but can also be found in asymptomatic persons.

Partial-thickness tears
Partial tears exhibit focal areas of mildly increased signal intensity on T1-weighted images, increasing signal intensity on T2-weighted images and contour irregularities (attenuated or thickened tendon), that is, grade 2 tendon. The signal abnormality does not extend throughout the entire thickness of the tendon. Free fluid may be present within the subacromial–subdeltoid bursa if the tear is located on the superior bursal surface or within the glenohumeral joint if the tear is located inferiorly at the articular surface.

Tendinitis
The differential diagnosis of acute or chronic tendinitis, degeneration and partial-thickness tears of the RC is difficult. The contrast behaviour on T1- versus T2-weighted images and the tendon morphology, such as thickness and thinning, are important in making the diagnosis. In acute tendinitis, MRI shows increased signal intensity on T1- and T2-weighted images without contour abnormalities or with only slight enlargement of the tendons. Chronic tendinitis may cause an increase in signal intensity on T1-weighted images without an increase in signal intensity on T2-weighted images or contour irregularities. The increase in signal intensity in both cases is moderate and results in an inhomogeneous appearance of the tendon.

MRA

MRA was introduced to overcome the limitations of standard MRI in diagnosing RC disease and shoulder instability.[13] Traditional arthrography is an X-ray procedure involving the intra-articular injection of radiopaque dye into the shoulder to demonstrate the anatomy of the joint by X-ray. The use of MRI as opposed to standard X-ray allows better visualisation of the soft tissues of the shoulder and is thought to extend the capabilities of conventional MRI because the contrast material delineates intra-articular structures and outlines abnormalities.[15] It has been claimed to improve

the differentiation and detection of partial RCTs and labral tears in comparison with standard MRI.[13] MRA can be conducted with different fluids, including pure saline and Ringer lactate, or with a mixture of saline and gadolinium contrast medium.

The relative benefits of these technologies for the diagnosis of shoulder pain in terms of accuracy and cost-effectiveness are currently unknown.

Treatment of shoulder pain

Most patients who present with shoulder pain in general practice will undergo conservative treatment based on patient history and clinical examination. A Dutch study of 335 patients with a new episode of soft tissue shoulder disorder found that in the subsequent year:

- 24% underwent a 'wait and see' policy or received non-steroidal anti-inflammatory drugs (NSAIDs)
- 29% were referred for physiotherapy
- 23% received a local injection of anaesthetic or steroid
- 19% were recommended physiotherapy and injections
- only 1% underwent surgery.[11]

In spite of treatment, only around 20% of patients report a complete recovery at 1 month[5,11] and 40–50% report that their symptoms have persisted or recurred 1 year after the initial consultation.[5,11,12]

In fact, evidence for the effectiveness of conservative treatments is limited. A recent systematic review[10] concluded that NSAIDs and subacromial glucocorticosteroid injections **may** be superior to placebo in improving the range of abduction in RC tendinitis but no conclusions could be drawn regarding treatment of adhesive capsulitis. Further systematic reviews have found inconclusive evidence for the efficacy of physiotherapy[18] or steroid injections[19] in the treatment of soft tissue shoulder disorders. No randomised controlled trials of surgical interventions have been identified.[10]

It is likely that the benefits of treatment (conservative or surgical) may vary for different underlying causes of shoulder pain,[6] lending further support to the need for accurate diagnosis at the earliest possible stage.

Reference tests used

A variety of invasive reference tests may be used, the most common being arthroscopy or arthrography.

Arthroscopy is an invasive procedure whereby a thin fibre-optic endoscope is introduced into the shoulder joint to allow direct visualisation of the internal structures. As discussed above, conventional arthrography is an X-ray procedure involving the intra-articular injection of radiopaque dye into the shoulder to demonstrate the anatomy of the joint. The single-contrast technique has been used to evaluate shoulder disorders since the 1930s.[20] The development of double-contrast arthrography (which includes the injection of 10–15 ml of room air in addition to the contrast agent) in the 1970s increased the ability of the technique to document suspected RCTs and allowed more accurate evaluation of the size of the defect and quality of the torn tendon edges.[13]

Both techniques have their proponents, but in general arthroscopy is recognised to allow direct visualisation of both sides of the RC and may be better at detecting partial-thickness tears. Its field of vision is nevertheless limited and may miss or misclassify some tears, for example on the articular surface of the RC tendon, or where the subacromial bursa is inflamed. Inter-observer variation in the criteria for diagnosis of tears between arthroscopists may also occur – terminology such as fraying, fibrillation, scuffing, fringe, degeneration or synovitis may be categorised as 'no tear' by some but as small or partial tears by others.

At one time, arthrography was thought to have high spatial resolution such that small and partial tears on the joint side of the cuff can be easily visualised. However, it is now recognised that arthrography lacks sensitivity.[21] In some tears the joint capsule acts as a ball valve and covers up the tear so it is not seen at the time of arthrography. Furthermore, as it relies on the passage of contrast material out of the joint space into the subacromial–subdeltoid bursa it may have difficulty depicting the more superficial and bursal aspects of the rotator cuff and may also leave partial tears undetected.

Open surgery has also been used as a reference test, although it potentially misses the joint surface so that articular surface tears or inferior surface tears may be missed. As open surgery often

involves a therapeutic in addition to a diagnostic component, the patients included in studies where it was used as a reference test may have more severe disorders warranting surgical intervention.

As RC abnormalities have also been documented on asymptomatic shoulders, use of invasive reference tests may not be useful. As a result, some authors claim that the subacromial injection test is a better reference standard, as it is better associated with clinical symptomatology of subacromial impingement syndrome. Others have used MRI despite the fact that its true accuracy is yet to be established.

Chapter 3
Methods

Search strategy

Literature was identified from several sources, including electronic databases and other sources including:

- general health and biomedical databases: MEDLINE, PubMED (current year), EMBASE, Science Citation Index, BIOSIS, AMED
- specialist electronic databases: DARE, EconLit
- research in progress: National Research Register (NRR), Current Controlled Trials, Clinical Trials
- checking of reference lists.

A comprehensive database of relevant articles was constructed. Literature was identified by combining search terms related to the shoulder with a methodological filter for identification of diagnostic test studies (see Appendix 1 for details). All databases were searched from 1985 to October 2001. Owing to time and resource constraints searches were restricted to English language only.

Inclusion criteria

The criteria for study inclusion in the systematic review were as follows.

Population

The majority of patients with shoulder pain are managed in general practice and it is important to focus on the concept of 'shoulder pain' rather than shoulder pathology. For the evaluation of clinical examination, therefore, all causes of shoulder pain were included.

For the evaluation of imaging techniques thought to be of particular benefit in the assessment of soft tissue structures, studies of adults with suspected soft tissue disorders of the shoulder including RC tendinitis or RCTs and shoulder bursitis were included. Studies that included only patients with shoulder pain resulting from other causes such as shoulder instability, arthritis or referred pain were excluded. Studies conducted in children and studies of cadavers were also excluded.

Setting

Studies conducted in any setting were included in the review, including general practice, accident and emergency and hospital clinics.

Interventions

Clinical examination and patient history

Any study that compared clinical examination and/or assessment of patient history in comparison with an acceptable gold standard for the evaluation of patients with shoulder pain was included.

Imaging tests

The following diagnostic imaging techniques in comparison with an acceptable reference standard or in comparison with each other were included in the review:

- ultrasound
- MRI
- MRA.

Evaluations of plain X-ray and CTA were excluded from the review, as these techniques are recognised to have limited value in the diagnosis of soft tissue lesions.

Studies using any reference test – arthroscopy, arthrography, surgery and in some cases MRI or the subacromial injection test – were included, and the appropriateness of the reference test used was assessed during quality assessment.

Outcome measures

The main focus of the review was the establishment of test accuracy, as this has been the focus of most primary studies of diagnostic tests. The outcomes assessed included detection of:

- clinical impingement syndrome
- RCT (full, partial or any).

Detection of labral disorders was excluded as these largely occur as a result of dislocation or instability.

At a minimum, accuracy studies had to report summary accuracy statistics or present sufficient raw data to allow these statistics to be calculated.

Studies that examined the effect of the included tests on diagnostic thinking, patient management or subsequent patient outcomes were also reviewed. This included any studies on patient preferences for the imaging techniques reviewed. Studies focusing on the establishment of technical efficacy alone were excluded.

Study design

All studies that compared a new test or strategy with an established reference test in patients **suspected of having** the target disorder were included. Studies, particularly case–control studies, which selected healthy control subjects were excluded.

For studies evaluating the impact of tests on patient management or patient outcomes, only prospective controlled studies were included.

Eligibility assessment

All studies were assessed for inclusion by two reviewers, and disagreements were resolved by consensus.

Quality assessment strategy

We proposed to appraise the methodological quality of included test accuracy studies using a formal quality assessment tool (Appendix 2) developed on behalf of the HTA programme.[22] The final tool was not available at the time of the data extraction, and therefore data on a series of potential quality items relating to the selection of the study cohort, performance of the reference standard and masked assessment of test results were extracted (Appendix 3). Only those items included in the final quality assessment tool are presented in this report.

Study quality was assessed independently by two reviewers. Any disagreements were resolved by consensus or if necessary by arbitration by a third reviewer.

Data extraction strategy

The extraction of study findings was conducted in duplicate using a pre-designed and piloted data extraction form to avoid any errors (Appendix 3). Any disagreements between reviewers was resolved by consensus or where necessary by arbitration by a third reviewer.

Methods of analysis

Test accuracy studies

Studies evaluating test accuracy were first grouped according to the index test evaluated, and sensitivity, specificity, positive and negative likelihood ratios (LRs) and 95% confidence intervals (CIs) were calculated for each test against the reference standard for each study. Tests with high sensitivity are most useful at ruling out disease, i.e. a negative result indicates that disease is not likely to be present. Tests with high specificity are most useful at ruling in disease, i.e. a positive result indicates that disease is likely to be present. LRs describe how many times a person with disease is more likely to receive a positive or negative test result than a person without disease. Generally, positive LRs of <10 cannot provide convincing evidence of the presence of disease, although those between 5 and 10 can give strong diagnostic evidence, depending on the pre-test probability and context to which they are applied.[23,24] Negative LRs of <0.1 are required to rule out disease convincingly on the basis of a negative test result; those between 0.2 and 0.1 may provide strong diagnostic evidence, again depending on the context.

Individual studies may have used varying explicit and implicit definitions for an abnormal result. This may be particularly true of imaging studies where interpretation of the same image can vary significantly between interpreters. Few studies reported explicit variations in cut-off, or presented enough information to allow data to be extracted to a standardised cut-off value or for a variety of cut-off points. Implicit variations in cut-off manifest themselves in a trade-off between sensitivity and specificity producing a correlation between true-positive and false-positive rates.[23,25] Spearman's rho was used to identify any such correlation.

Where no correlation was revealed, and studies were otherwise sufficiently homogeneous, pooled estimates of sensitivity and specificity were calculated. Summary estimates were produced with 95% CIs. Random effects methods for meta-analysis were used, as heterogeneity between test statistics is routinely encountered in diagnostic meta-analyses.

Subgroup analyses

Potential sources of heterogeneity were investigated by conducting subgroup analyses according to features of the population (spectrum), test and reference test and study quality.

Studies evaluating outcomes other than test accuracy

For studies comparing patient outcomes, diagnostic and therapeutic impact comparisons or the cost-effectiveness of different tests, a narrative synthesis was undertaken.

Chapter 4
Results

Quantity of research available

No existing systematic reviews evaluating the accuracy of diagnostic tests for the investigation of shoulder pain were identified. One systematic review of the effect of MRI of the shoulder on patient outcomes[26] and one cost-effectiveness analysis of arthrography versus MRI[27] were identified. In addition, several systematic reviews on the effectiveness of treatment of shoulder pain were found.[10,18,19,28,29]

The titles and abstracts of 1515 papers were screened for eligible studies, the full papers of 179 papers were retrieved for more detailed evaluation and 73 studies of test accuracy published in 74 separate publications were included (see *Figure 1* for a flowchart of the screening process). The reasons for exclusion of 98 studies are provided in Appendix 4. Ten studies evaluated more than one index test, but they only presented data for each test against the reference standard; none of them presented comparative data for each test against the other.

Clinical examination

Description and quality of included studies

Ten cohort studies were included in the review.[30–39] Summary details of the methods and quality assessment results of these studies are provided in *Table 2* and full details in Appendix 5.

Interventions

Seven studies[30–33,35,38,39] examined the accuracy of individual clinical examination tests and six[30,33–37] estimated the accuracy of clinical examination *per se* or the combination of two or more positive test results. A total of 23 different signs, symptoms or clinical examination tests were used throughout the studies but only four were examined in more than one study:

- Jobe test[31,32]
- Neer test[30,35]
- Hawkins test[30,35]
- painful arc.[30,33]

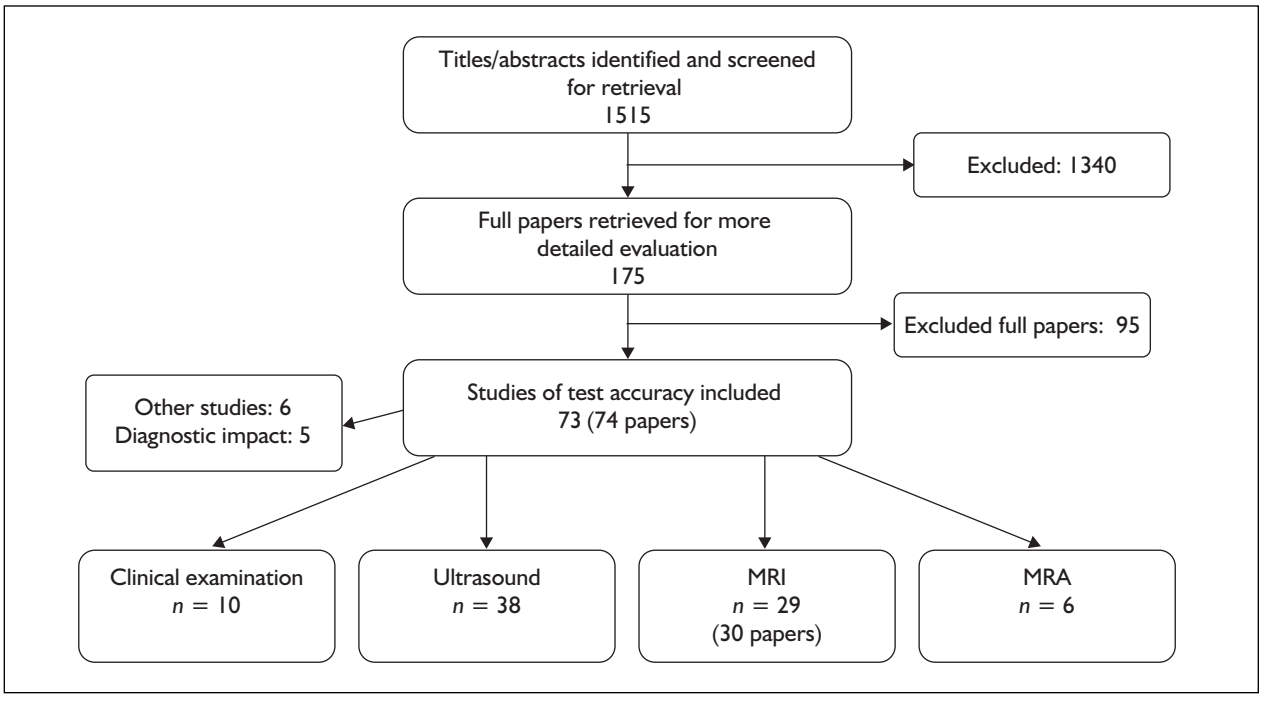

FIGURE 1 *Flowchart of screening process*

TABLE 2 *Clinical examination: study methods and quality assessment*

Study	Design[a]	Sample size	Mean age (years)	% males in the sample	Test details	Ref[b]	Prevalence[c]	Appropriate spectrum[d]	Eligibility criteria stated[d]	Appropriate ref. test[d]	Disease progression bias[d]	Partial verification[d]	Differential verification[d]	Incorporation bias[d]	Details of index test[d]	Details of ref. test[d]	Test review bias[d]	Diagnostic review bias[d]	Clinical info. available[d]	Indeterminate missing[d]	Withdrawal bias[d]
Calis, 2000[30]	P1	125	52	40	Single tests + combined	SIT	70[I]	Y	Y	?	nr	N	N	N	Y	Y	N	?	?	N	N
Hertel, 1996[31]	P3	100	51	74	Single tests	S/AS	72	?	Y	Y	nr	N	Y	N	Y	N	N	?	?	N	N
Itoi, 1999[32]	P1	143	43	77	Single tests	M	24[F]	?	N	?	nr	N	N	N	Y	Y	N	?	?	N	N
Litaker, 2000[33]	R	448	57	63	Single tests + expert diagnostics	AG	67	?	Y	N	nr	?	N	N	Y	N	?	?	?	N	N
Lyons, 1992[34]	R	42	?	60	Clin. exam.	S	81	?	N	Y	nr	N	N	N	Y	N	?	?	?	N	?
MacDonald, 2000[35]	P2	85	40	73	Single tests + combined	AS	28	?	N	Y	nr	N	N	N	Y	N	N	?	?	N	N
Read, 1998[36]	?	42	44	?	Clin. exam.	AS/S	81	?	Y	Y	Y	N	Y	N	Y	N	?	?	?	N	N
Wnorowski, 1997[37]	R	31	30	?	Clin. exam.	AS/AG	35	?	N	Y	Y	N	Y	N	N[e]	?	?	?	?	N	N
Wolf, 2001[38]	R	109	51	61	Single test	AS	42[F]	?	?	Y	nr	N	N	N	Y	N	?	?	?	N	N
Zaslav, 2001[39]	P2	110	44	59	Single test	AS	24[N]	N	Y	Y	nr	N	N	N	Y	N	N	?	?	N	N

[a] Design: P1, prospective, unselected sample of patients suspected of RC disorder; P2, prospective, all patients referred for reference test; P3, prospective, selected patients underwent reference test; R, retrospective; ?, design not reported.
[b] Reference test: AG, arthrography; AS, arthroscopy; S, surgery; M, MRI; SIT, subacromial injection test.
[c] Prevalence (%) of any tear, unless specified: F, full-thickness tear; I, impingement syndrome; N, non-outlet impingement.
[d] Responses to quality assessment criteria: N, No; Y, yes; ?, cannot tell; nr, not reported.
[e] Applies only to clinical examination – details of MRI were provided.

Only one study did not provide sufficient details of the tests used, reporting only that 'clinical diagnosis' was evaluated.[37] The remaining studies all scored 'yes' on this item.

Outcomes

The majority of studies evaluated the ability of clinical examination to identify patients with RCTs[31,33–35,37] or, in two cases, specifically full-thickness RCTs.[32,38] Two other studies used clinical examination to differentiate impingement syndrome from other causes of shoulder pain,[30,36] and another[39] 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.

Sample details

Studies were of variable size, although six included at least 100 participants. The mean sample size was 123. None were conducted in a primary care setting but were largely conducted in orthopaedic or specialist shoulder units, sometimes in association with radiology departments. One study was conducted in a physical medicine and rehabilitation clinic to which patients had been referred from rheumatology or orthopaedics, but could also have been self-referred.[30] As a result, these studies are not likely to provide an indication of how well clinical examination would perform on a relatively unselected population. In fact, only one was judged to have included an appropriate spectrum of patients: although Calis and colleagues[30] did not report the mean duration of shoulder pain, it was a prospective study recruiting consecutive patients referred to a physical medicine and rehabilitation clinic.

Eight of the other nine studies included only patients already selected to undergo the reference test[31,33–39] (which in three cases included open surgery, i.e. a therapeutic rather than diagnostic intervention such that they included patients already deemed to require surgical intervention) and although six reported recruitment of consecutive patients,[31–33,35,38,39] five did not report sufficient patient characteristics to allow a judgement of spectrum to be made, and in the sixth[39] participants were reported to have suffered shoulder pain for a mean of 10.9 months.

The prevalence of RC disorders was therefore high in most studies (overall mean 61%), although it varied according to the setting and outcome used. For the seven studies looking to detect the presence of impingement syndrome[30,35,36] or any RCT,[31,33–35,37] prevalence ranged from 28 to 81%, although in five cases prevalence was >50%. In the two studies looking to detect the presence of full-thickness tears, prevalence was 24%[32] and 42%.[38] The overall mean age of included patients was 46 years. In most studies the mean age lay between 40 and 52 years, but in one the median age was only 30 years[37] and in another 57.4 years.[33] Older patients are more likely to be suffering from 'classical' outlet impingement rather than impingement due to labral lesions. Two studies did not report the sex of included patients, but of the others, all but one[30] included a majority of male patients (overall mean 63%).

Reference test

In six of the studies the reference standard used was likely to have correctly classified the target condition,[31,34–37,39] although in two of these[36,37] the mean delay between application of the index and reference tests was at least 2 months such that resolution or advancement of the severity of disorder was possible. Two of the remaining studies used either the subacromial injection test (SIT)[30] or MRI[32] as the reference test, the accuracies of which have yet to be established.

Partial verification was not present in nine studies. In the other,[33] 501 eligible patients were identified, but complete data were available for only 448. Information was not provided regarding the missing data, but it is possible that some did not undergo the reference test, so this study was scored as 'unclear'. Differential verification was used in three of the studies. Two used either arthroscopic or open surgery,[31,36] one of which[36] clearly stated that open surgery was used if a full RCT was suspected and arthroscopy used for other cases. A further study[37] used arthroscopy with or without arthrography according to the result of the MRI investigation. Only two studies provided sufficient details of the reference test used, that is, those using the SIT[30] and MRI[32] as their reference standard. One other[37] provided some details of the arthrographic procedure performed but not of arthroscopy.

Test interpretation

The subjectivity of clinical examination makes blinding an important issue. None of the 10 studies reported that the final diagnosis reached from clinical examination was attained without knowledge of the reference test result; however, as clinical examination would always have been

performed first, it is likely, at least in the prospective studies,[30–32,35,39] that the test result was not influenced by the reference test result. In the remainder, where retrospective chart review was used to establish both diagnoses, it is possible that test review bias was present. Diagnostic review bias could have influenced the establishment of the reference diagnosis in all 10 studies. As knowledge of age, sex and symptoms is an integral part of clinical examination, such information should have been available to all those performing the clinical examination tests. None of the studies reported that any blinding to such details was performed, so all scored as unclear; however, it does seem reasonable to assume that no such blinding was performed.

Few studies reported details of those interpreting the tests other than that they were orthopaedists or radiologists, perhaps specialising in shoulder disorders. None appeared to be physiotherapists or other healthcare practitioners who might encounter people with shoulder pain in routine practice.

Indeterminate results were never mentioned in relation to clinical examination, so all studies scored 'no' on this item. It does seem possible, however, that indeterminate results might have occurred as it may not always possible to make a definite diagnosis on the basis of clinical examination alone. However, as results were always reported for the whole sample, either indeterminate results were excluded before the sample was selected (e.g. in retrospective studies) or were simply recorded as positive or negative results.

Withdrawal bias did not appear to be present in nine studies. In one of these,[33] 53/501 patients were not included owing to incomplete data, but they were reported not to differ significantly from included patients in terms of demographic characteristics or historical or physical examination features. In another,[39] five patients were discovered not to meet the eligibility criteria. In the tenth study,[34] two patients' case notes were lost and a third was excluded as no estimate of the size of RC was provided; withdrawal bias may have been present.

Results

Individual clinical examination tests

The tests evaluated tended to be either highly sensitive or highly specific, and very few demonstrated both high sensitivity and specificity (*Table 3*). As a result, few tests provided convincing evidence of the presence or absence of disease in the settings in which they were applied. Individual tests did perform particularly well in the study by Hertel and colleagues:[31] the external rotation lag sign, the drop test, lift-off test and internal rotation lag sign all had positive LRs of >10 (although the sample size was small and CIs were very wide). The internal rotation lag sign also had a very low negative LR of 0.0 (95% CI: 0.0 to 0.2). Other tests demonstrating high positive and negative LRs were the Rent test[38] and internal rotation resistance strength test.[39]

Of the remaining tests, several produced sensitivities of >80% for the detection of impingement syndrome or RCT, including the Neer test,[30,35] Hawkins test,[30,35] the horizontal adduction test,[30] the Jobe test,[31,32] impingement sign and arc of pain.[33] Tests with specificities of >80% included drop arm test,[30] Yergason,[30] Speed test,[30] and passive external rotation.[33] In short, individual tests are either good at ruling out RC disorders when negative (high sensitivity) or at ruling in such disorders when positive (high specificity), but small sample sizes mean that there is not really conclusive evidence for any single test that can conclusively diagnose RC disorders.

Combination of tests or 'clinical examination' per se

When two or more tests are used in combination, the resulting positive LRs are all <5, indicating that a positive diagnosis on the basis of clinical examination as a whole is not a convincing result (*Table 4*). However, in four studies[33–36] negative LRs were sufficiently low to confirm that disease is absent in those with a negative diagnosis.

This was further investigated by pooling the results of the four studies that evaluated 'clinical examination' *per se*, that is, without specific combinations of results of two or more tests. The resulting forest plots of sensitivity and specificity are presented in *Figure 2*. These demonstrate that sensitivities and specificities are relatively homogeneous (almost all of the respective CIs are overlapping), although the estimate of sensitivity from Wnorowski and colleagues[37] is outlying. Spearman's rho was 0.80, which is not statistically significant ($p = 0.2$). The plot of sensitivity against the false-positive rate confirms that the studies are generally clustered together (again with the Wnorowski study outlying) with no evidence for a threshold effect (*Figure 3*).

For sensitivity, the chi-squared test for heterogeneity was 11.8 on three degrees of

TABLE 3 *Clinical examination: accuracy of individual clinical examination tests*

Study	Pr[a] (%)	Test description	Se[b] (95% CI)	Sp[b] (95% CI)	LR+[b] (95% CI)	LR–[b] (95% CI)
Calis, 2000[30]	70[I]	Neer	0.89 (0.80 to 0.94)	0.32 (0.19 to 0.47)	1.3 (1.0 to 1.6)	0.4 (0.2 to 0.8)
		Hawkins	0.92 (0.84 to 0.96)	0.24 (0.13 to 0.39)	1.2 (1.0 to 1.5)	0.3 (0.2 to 0.9)
		Horizontal adduction test	0.82 (0.72 to 0.88)	0.29 (0.17 to 0.45)	1.1 (0.9 to 1.4)	0.6 (0.3 to 1.2)
		Painful arc	0.69 (0.59 to 0.78)	0.55 (0.40 to 0.70)	1.5 (1.0 to 2.3)	0.6 (0.4 to 0.9)
		Drop arm	0.37 (0.27 to 0.47)	0.87 (0.73 to 0.94)	2.8 (1.2 to 6.6)	0.7 (0.6 to 0.9)
		Yergason	0.32 (0.23 to 0.43)	0.82 (0.67 to 0.91)	1.7 (0.8 to 3.6)	0.8 (0.7 to 1.0)
		Speed	0.8 (0.4 to 0.16)	0.97 (0.87 to 1.00)	3.1 (0.4 to 24)	0.9 (0.9 to 1.0)
Hertel, 1996[31]	72[A c] (n = 87)	Jobe	0.84 (0.73 to 0.91)	0.58 (0.39 to 0.76)	2.0 (1.2 to 3.3)	0.3 (0.1 to 0.5)
		External rotation lag sign (ERLS)	0.70 (0.58 to 0.80)	1.00 (0.86 to 1.00)	34.8 (2.2 to 543.1)	0.3 (0.2 to 0.4)
		Drop test	0.21 (0.12 to 0.32)	1.00 (0.86 to 1.00)	10.5 (0.7 to 170.8)	0.8 (0.7 to 0.9)
	55[A c] (n = 53)	Lift-off	0.62 (0.44 to 0.77)	1.00 (0.86 to 1.00)	30.8 (2.0 to 486.4)	0.4 (0.2 to 0.6)
		Internal rotation lag sign (IRLS)	0.97 (0.83 to 0.99)	0.96 (0.80 to 0.99)	23.2 (3.4 to 158.1)	0.0 (0.0 to 0.2)
Itoi, 1999[32]	24[F]	Full can – pain	0.66 (0.49 to 0.79)	0.64 (0.54 to 0.72)	1.8 (1.3 to 2.6)	0.5 (0.3 to 0.9)
		Full can – muscle weakness	0.77 (0.61 to 0.88)	0.74 (0.65 to 0.81)	3.0 (2.1 to 4.3)	0.3 (0.2 to 0.6)
		Full can – pain, muscle weakness or both	0.86 (0.71 to 0.94)	0.57 (0.48 to 0.66)	2.0 (1.6 to 2.6)	0.2 (0.1 to 0.6)
		Empty can (Jobe) – pain	0.63 (0.46 to 0.77)	0.55 (0.45 to 0.64)	1.4 (1.0 to 1.9)	0.7 (0.4 to 1.1)
		Empty can (Jobe) – muscle weakness	0.77 (0.61 to 0.88)	0.68 (0.58 to 0.76)	2.4 (1.7 to 3.3)	0.3 (0.2 to 0.6)
		Empty can (Jobe) – pain, muscle weakness or both[d]	0.89 (0.74 to 0.95)	0.50 (0.41 to 0.59)	1.8 (1.4 to 2.2)	0.2 (0.1 to 0.6)
Litaker, 2000[33]	67[A]	Night pain	0.88 (0.84 to 0.91)	0.20 (0.14 to 0.27)	1.1 (1.0 to 1.2)	0.6 (0.4 to 1.0)
		Muscle atrophy – supraspinatus	0.55 (0.50 to 0.61)	0.73 (0.65 to 0.79)	2.0 (1.5 to 2.7)	0.6 (0.5 to 0.7)
		Muscle atrophy – infraspinatus	0.55 (0.50 to 0.61)	0.73 (0.65 to 0.79)	2.0 (1.5 to 2.7)	0.6 (0.5 to 0.7)
		Passive elevation	0.30 (0.25 to 0.36)	0.78 (0.71 to 0.84)	1.4 (1.0 to 2.0)	0.9 (0.8 to 1.0)

continued

TABLE 3 *Clinical examination: accuracy of individual clinical examination tests (cont'd)*

Study	Pr[a]	Test description	Se[b] (95% CI)	Sp[b] (95% CI)	LR+[b] (95% CI)	LR–[b] (95% CI)
		Passive external rotation	0.19 (0.15 to 0.24)	0.84 (0.77 to 0.89)	1.2 (0.8 to 1.8)	1.0 (0.9 to 1.1)
		Impingement sign	0.97 (0.95 to 0.99)	0.9 (0.5 to 0.15)	1.1 (1.0 to 1.1)	0.3 (0.1 to 0.7)
		Weakness with elevation	0.64 (0.58 to 0.69)	0.65 (0.57 to 0.73)	1.8 (1.4 to 2.3)	0.6 (0.5 to 0.7)
		Weakness with external rotation	0.76 (0.71 to 0.80)	0.57 (0.49 to 0.65)	1.8 (1.5 to 2.2)	0.4 (0.3 to 0.5)
		Arc of pain	0.97 (0.95 to 0.99)	0.10 (0.6 to 0.16)	1.1 (1.0 to 1.1)	0.3 (0.1 to 0.6)
		History of trauma	0.36 (0.31 to 0.41)	0.73 (0.65 to 0.79)	1.3 (1.0 to 1.8)	0.9 (0.8 to 1.0)
MacDonald, 2000[35]	28[A]	Neer	0.88 (0.69 to 0.96)	0.43 (0.31 to 0.55)	1.5 (1.2 to 2.0)	0.3 (0.1 to 0.9)
		Hawkins	0.83 (0.64 to 0.93)	0.51 (0.39 to 0.63)	1.7 (1.2 to 2.3)	0.3 (0.1 to 0.8)
	28[B]	Neer	0.75 (0.55 to 0.88)	0.48 (0.36 to 0.60)	1.4 (1.0 to 2.0)	0.5 (0.3 to 1.1)
		Hawkins	0.92 (0.74 to 0.98)	0.44 (0.33 to 0.57)	1.6 (1.3 to 2.1)	0.2 (0.1 to 0.4)
	53[I]	Neer	0.89 (0.77 to 0.95)	0.60 (0.45 to 0.74)	2.1 (1.3 to 3.2)	0.4 (0.2 to 0.6)
		Hawkins	0.77 (0.64 to 0.87)	0.63 (0.47 to 0.76)	2.2 (1.5 to 3.3)	0.2 (0.0 to 0.7)
Wolf, 2001[38]	42[F]	Rent test	0.96 (0.85 to 0.99)	0.97 (0.89 to 0.99)	30.1 (7.7 to 118.0)	0.0 (0.0 to 0.2)
Zaslav, 2001[39]	24[N]	Internal rotation resistance strength test (IRRST)	0.88 (0.71 to 0.96)	96 (0.90 to 0.99)	24.8 (8.1 to 75.9)	0.1 (0.0 to 0.3)

[a] Prevalence (%) of: A, any tear; F, full-thickness tear; I, impingement syndrome; B, bursitis; N, non-outlet impingement.
[b] Se, sensitivity; Sp, specificity; LR+, positive likelihood ratio; LR–, negative likelihood ratio.
[c] Results reported separately for those with supraspinatus and subscapularis defects. First three tests evaluated supraspinatus, last two tests evaluated subscapularis.
[d] These are the usual criteria for interpretation of the Jobe test.

freedom (df) ($p = 0.008$), indicating the presence of significant heterogeneity (although with so few studies, the test has low power). The estimate of pooled sensitivity was 0.90 (95% CI: 0.87 to 0.93). Again, only the Wnorowski study does not include the overall pooled value within its CI. The clinical examination technique used in this study was not reported; however, the sample size was very small (31 patients), patients were considerably younger (median age 30 years), and tear prevalence was very low in comparison with the others (35% versus an overall prevalence across the four studies of 67%). When this study is removed, there is no significant heterogeneity $\chi^2 = 2.2$, $p = 0.3$, and pooled sensitivity becomes 0.91 (95% CI: 0.88 to 0.94).

For specificity estimates, the chi-squared test for heterogeneity among all four studies was 1.9 on df = 3 ($p = 0.6$), indicating no significant heterogeneity. The estimate of pooled specificity was 0.54 (95% CI: 0.47 to 0.61).

Combining LRs using a Mantel–Haenszel fixed effects model gives pooled positive and negative LRs (95% CI) of 1.95 (1.65 to 2.31) and 0.21 (0.15 to 0.28) (*Figure 4*). As with the sensitivities there is significant heterogeneity in negative LRs between

TABLE 4 *Clinical examination: accuracy of clinical examination* per se *or combinations of two or more individual tests*

Study	Pr[a]	Test description	Se[b] (95% CI)	Sp[b] (95% CI)	LR+[b] (95% CI)	LR–[b] (95% CI)
Calis, 1997[30]	70[I]	⩾3 positive tests out of total of 7	0.85 (0.76 to 0.91)	0.45 (0.30 to 0.60)	1.5 (1.1 to 2.1)	0.3 (0.2 to 0.6)
		⩾4 positive tests out of total of 7	0.70 (0.60 to 0.79)	0.66 (0.50 to 0.79)	2.0 (1.3 to 3.2)	0.5 (0.3 to 0.7)
		⩾5 positive tests out of total of 7	0.39 (0.29 to 0.50)	0.87 (0.73 to 0.94)	3.0 (1.3 to 7.0)	0.7 (0.6 to 0.9)
		⩾6 positive tests out of total of 7	0.30 (0.21 to 0.40)	0.89 (0.76 to 0.96)	2.8 (1.1 to 7.6)	0.8 (0.7 to 0.9)
		All 7 tests positive	0.5 (0.2 to 0.11)	0.97 (0.87 to 1.00)	1.7 (0.2 to 15.1)	1.0 (0.9 to 1.1)
Litaker, 2000[33]	67[A]	Expert diagnosis (10 tests performed plus radiography)[c]	0.90 (0.87 to 0.93)	0.53 (0.45 to 0.61)	1.9 (1.6 to 2.3)	0.2 (0.1 to 0.3)
Lyons, 1992[34]	81[A]	Clinical examination (3 tests performed)[c]	0.91 (0.77 to 0.97)	0.75 (0.41 to 0.93)	3.6 (1.1 to 12.1)	0.1 (0.0 to 0.4)
MacDonald, 2000[35]	28[A]	Neer and Hawkins positive	0.83 (0.64 to 0.93)	0.56 (0.43 to 0.67)	1.9 (1.4 to 2.6)	0.3 (0.1 to 0.7)
		Neer or Hawkins positive[c]	0.88 (0.69 to 0.96)	0.38 (0.27 to 0.50)	1.4 (1.1 to 1.8)	0.3 (0.1 to 1.0)
	28[B]	Neer and Hawkins positive	0.71 (0.51 to 0.85)	0.51 (0.39 to 0.63)	1.4 (1.0 to 2.1)	0.6 (0.3 to 1.1)
		Neer or Hawkins positive	0.96 (0.80 to 0.99)	0.41 (0.30 to 0.54)	1.6 (1.3 to 2.0)	0.1 (0.0 to 0.7)
Read, 1998[36]	81[I]	Clinical examination (~6 signs/symptoms)	0.97 (0.85 to 0.99)	0.63 (0.31 to 0.86)	2.6 (1.1 to 6.3)	0.0 (0.0 to 0.4)
Wnorowski, 1997[37]	35[A]	Clinical examination (no details)[c]	0.55 (0.28 to 0.79)	0.50 (0.30 to 0.70)	1.1 (0.5 to 2.2)	0.9 (0.4 to 2.0)

[a] Prevalence of: A, any tear; I, impingement syndrome; B, bursitis.
[b] Se, sensitivity; Sp, specificity; LR+, positive likelihood ratio; LR–, negative likelihood ratio.
[c] Indicates those results used in meta-analysis.

studies ($\chi^2 = 17.22$, df = 3, $p = 0.001$). Using a DerSimonian and Laird random effects model to incorporate the between study variation gives pooled positive and negative LRs (95% CI) of 1.88 (1.34 to 2.65) and 0.21 (0.07 to 0.61), respectively. When the Wnorowski study is again removed, no heterogeneity remains and the pooled negative LR (95% CI) is 0.16 (0.11 to 0.24).

Litaker and colleagues[33] used a stepwise logistic regression model that included those signs, symptoms and test results associated with the presence of an RCT to develop a scoring system in order to find the best clinical cutpoint using LRs. Three factors were found to have a significant ($p < 0.05$) association with RCT: weakness on external rotation, age ⩾65 years and experience of night pain. These three factors were assigned scores of 2, 2 and 1, respectively, and a summary score of 4 or more was found best to differentiate patients with and without abnormal arthrograms. The associated positive LR for this probability score was 9.84, compared with only 1.93 for the expert diagnosis. Bearing in mind the limitations of conclusions drawn from a single study, this suggests that there may be specific factors that, when found in combination, are more accurate at detecting RCTs than diagnoses drawn from a more general clinical examination.

Discussion

Too few studies evaluated the same test to draw any conclusions regarding individual tests.

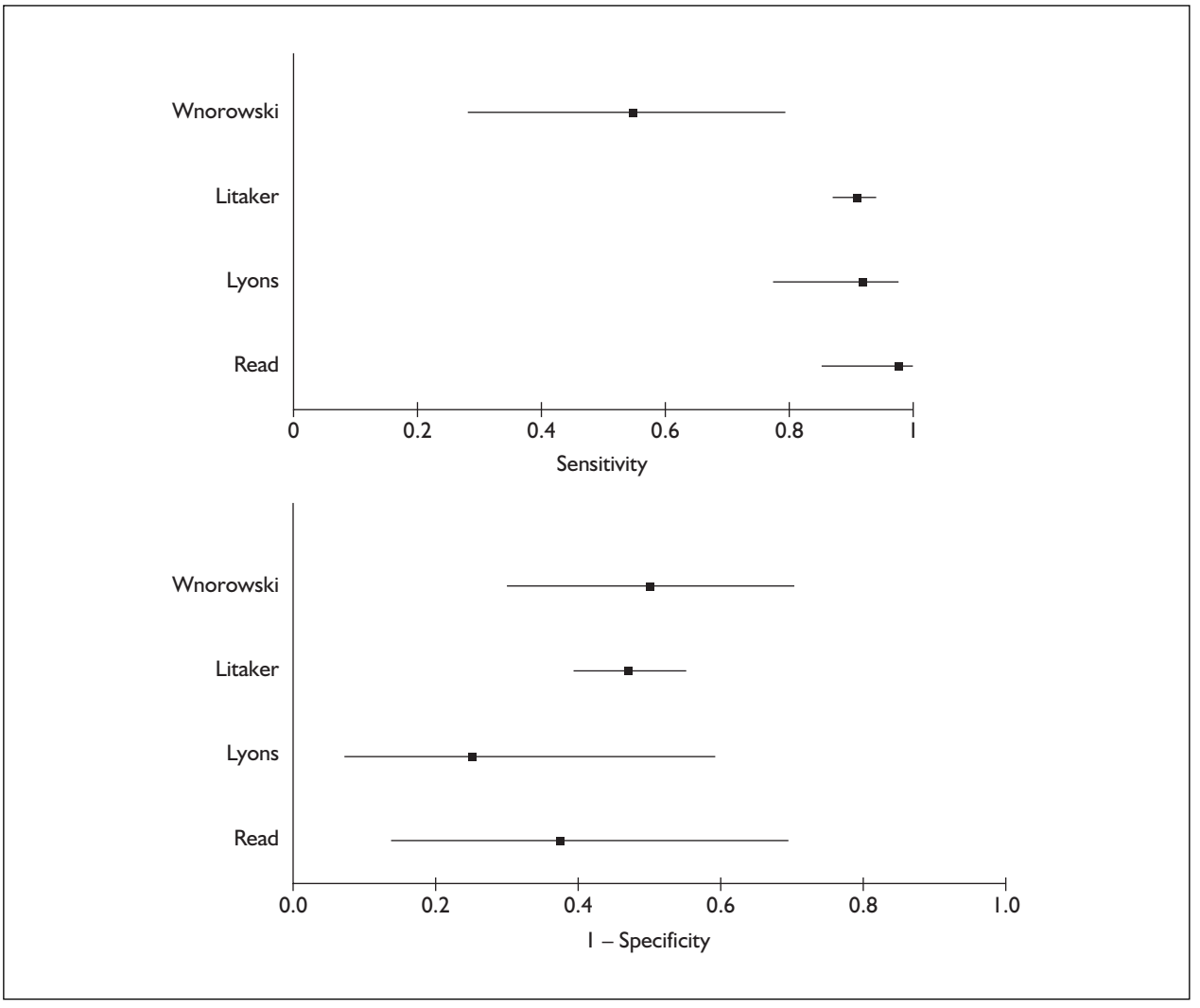

FIGURE 2 *Clinical examination: sensitivity and false-positive rates (1 – specificity). Detection of impingement syndrome or any RCT*

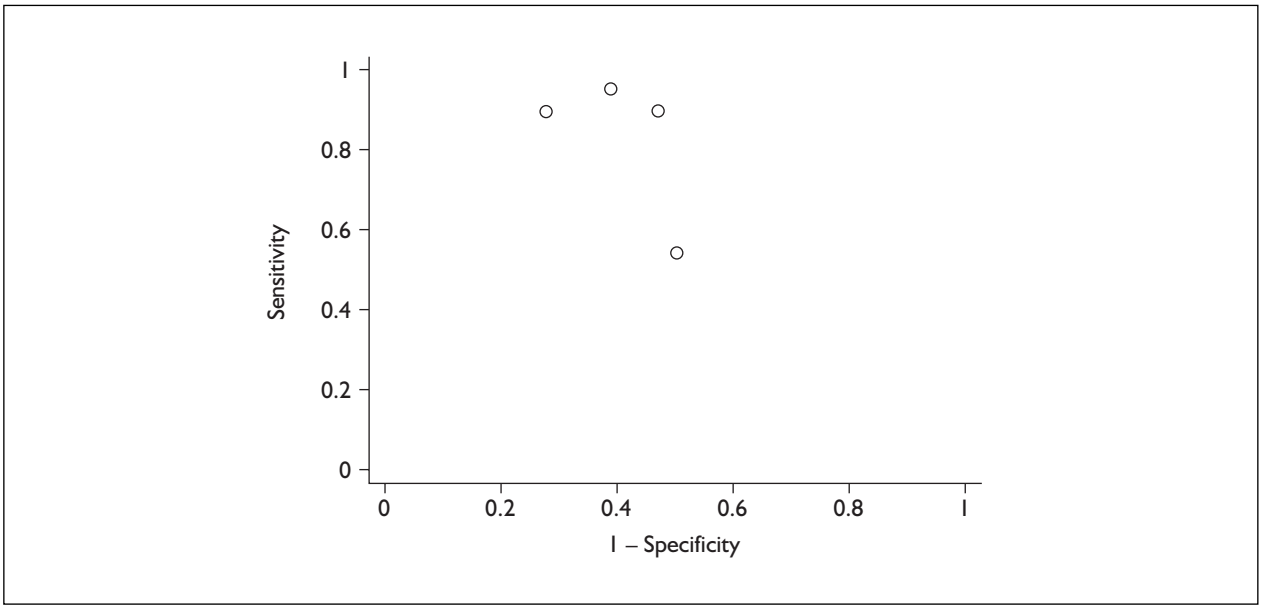

FIGURE 3 *Clinical examination: ROC plot of sensitivity versus 1 – specificity*

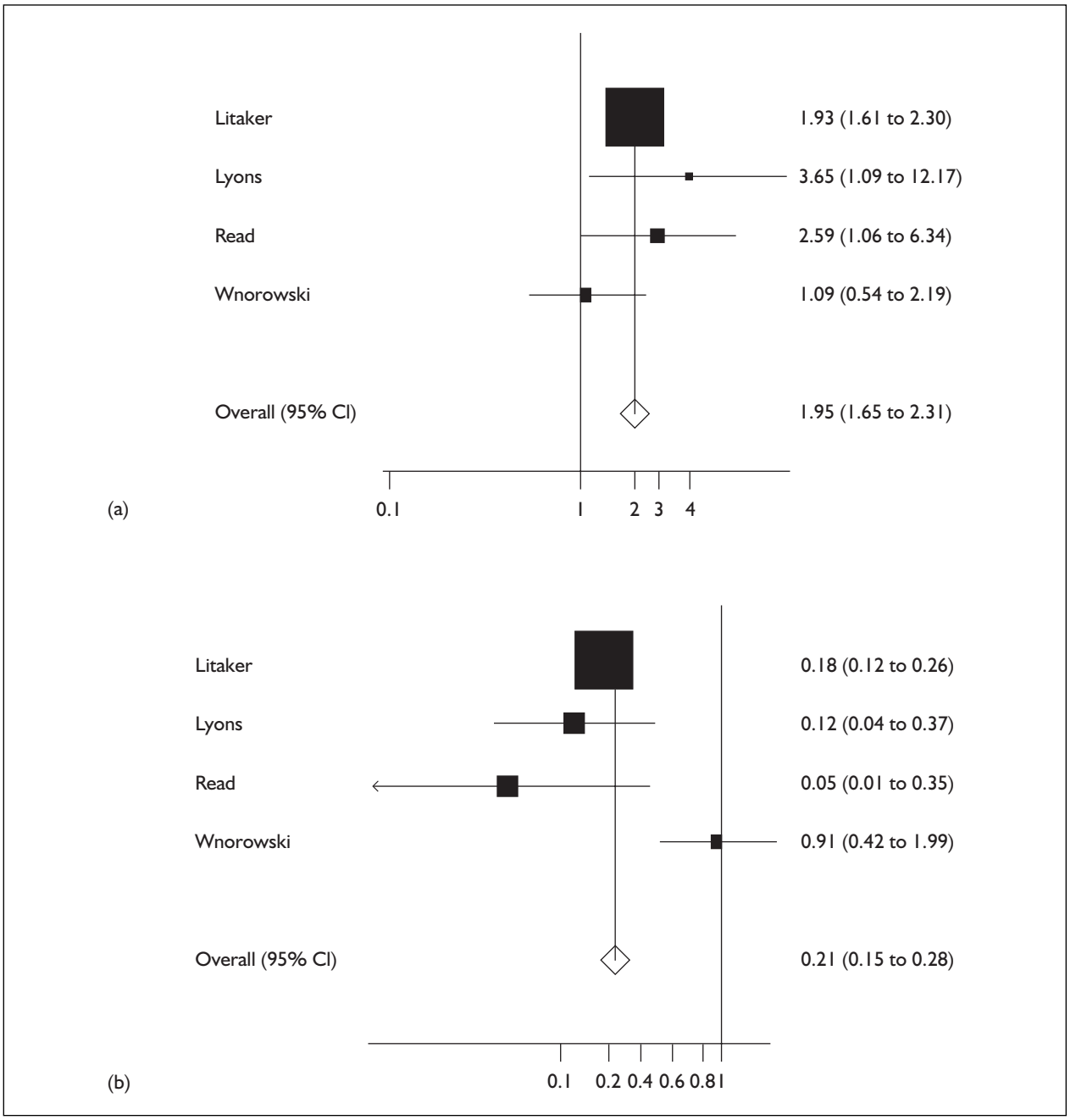

FIGURE 4 *Clinical examination: (a) positive and (b) negative LR*

Although the number of studies included was small, the meta-analysis confirms that clinical examination may be useful at ruling out RC disorders (high sensitivity and negative LR) but less accurate at detecting such disorders when they are present (low specificity and positive LR). Too few studies were available and studies were too poorly reported to allow any investigation of particular aspects of clinical examination or methodologies employed in these studies. In addition, the problem of selection bias from studies which used surgery as the definitive test needs to be borne in mind. Such studies will tend to include patients with more severe shoulder disorders and may not give results which are generalisable to the cases seen in primary care or sports medicine clinics.

Ultrasound

Description and quality of included studies

A total of 38 cohort studies investigating the accuracy of ultrasound were identified. Summary details of the methods and results are

provided in *Tables 5–8* and fuller details in Appendices 6–8.

Interventions

There was a lot of variation in the frequencies of the ultrasound transducers used. Although frequencies of >10 MHz are common in practice today, most studies used only 5 MHz (six studies) or 7 MHz (19 studies) or a combination of the two (six studies); only three studies exclusively employed transducers of 10 MHz or over.[40–42] The techniques used were described in some detail (in all except three studies[43–45]), and were largely based on those of Mack and colleagues,[46,47] Middleton and colleagues[41,48] and Crass and colleagues.[40,49]

Outcomes

Studies generally concentrated on the detection of 'any' RCT (27 studies), although 17 studies also attempted to differentiate full or partial tears using the technique (see Appendix 7).

Sample details

The 38 studies included a total of 2435 patients (1270 cases and 1165 controls) with a mean overall prevalence of RCT of 52%. Sample sizes were generally small, with a mean of 64 and only nine studies including at least 100 participants.[40,41,50–56] Around half (17) were prospective in design, five were retrospective and the remainder did not give details of the study design. The study setting was not always reported, but where it was, only two[57,58] were conducted in sports medicine or physical medicine centres rather than radiology or orthopaedics departments. Only 17 studies gave more than a general indication of the eligibility criteria used to include patients. One small study included mainly patients with rheumatoid arthritis[59] and four explicitly excluded those with arthritis[60,61] or other underlying inflammatory disease.[62,63] Surprisingly limited details of the samples were provided: 11 did not report mean age, 10 did not report the sex distribution of included participants and only a handful provided information on characteristics other than age and sex. The mean age of included participants across the studies was 51 years. For 20 studies the mean age was in the range 45–55 years; three studies included on average younger patients;[36,61,64] and four included older patients.[55,56,60,65] Where it was reported, most included a majority of male patients (all except seven studies[44,55,56,59,60,66,67]), overall mean 59% male.

Eleven studies provided sufficient information on which to judge the spectrum of included patients, but only one[68] was judged to have included an appropriate spectrum of patients. In the others, patients were reported to have experienced pain for at least 6 months or more[55,58,60–64,69] (in three cases for a mean of 2 years[62,63,69]) or specifically stated that only problem patients[70] or those with persistent symptoms[56] were included. The remaining studies did not provide sufficient information on which to make a judgement; however given that all except two[71,72] only included patients undergoing the reference test, it is unlikely that they were broadly representative of patients encountered in practice. This is also reflected in the prevalence of RCTs in the study samples as mentioned above; only nine studies had prevalences of <40%.[41,43,52,64,65,68,71,73,74]

Reference test

One-third of the studies used only arthrography as the reference test and were judged not to have employed a suitable reference standard.[41,43,45,47,50,57,59,65,68,71–73,75] In a further four,[67,70,74,76] arthrography was used as the reference standard in at least some of the patients enrolled – the reference test in these studies was scored as unclear. In all of the remaining studies except one using MRI,[60] the reference standard used was likely to have correctly classified the target condition. Sufficient details to allow replication of the reference tests was provided in only 16.[41,43,44,54,56,59–63,69,71,74,75,77,78]

Partial verification bias was also a problem, being clearly present in 10 studies[40,45,52,58,66,67,70,74,76,79] where only a subset of those undergoing ultrasound actually underwent the reference test, and clearly not present in only nine.[47,51,57,60,64,71–73,78] In the other 19 studies, this item was scored as unclear as it appeared that only those who underwent the reference test were included in the study; others who may have undergone ultrasound or been eligible to undergo ultrasound were not reported. Differential verification, where more than one reference test was used, was present in 16 studies,[36,42,47,53,54,56–58,62,64,66,67,70,74,76,79] although in only one of these[56] was it explicitly stated that the choice of reference test was based on the ultrasonographic classification.

Nine studies reported that ultrasound and the reference test were performed on the same day or within 1 month of each other,[41,45,60,63,70–72,74,75] and five reported a mean/median interval between tests of >1 month.[36,53,55,78,79] In one of these,[53] the authors reported that one of the tears missed by ultrasound was very large but was not confirmed by surgery until 5 months after

ultrasound had been performed. In the remaining studies, the delay between the tests was either not clearly reported (e.g. where interval was reported only to be <6 months but actual mean/median delay was not given[66,67]) or was not reported.

Test interpretation

Twenty-three studies explicitly reported that the diagnosis from ultrasound was reached without knowledge of the reference test result or was made and recorded during real-time imaging (*Table 5*). As the ultrasound examination would almost always be performed before the reference test, it may be safe to assume that those coded as unclear in the quality assessment were in fact blinded to the reference test results. In one study[72] the ultrasound diagnosis for one patient was deemed inconclusive and was not made until the results of the reference test were known. Diagnostic review bias (knowledge of the ultrasound result) was not present in 11 studies, but could have affected results in a further five;[36,55,64,67,74] the remainder did not state whether the ultrasound results were available to the person conducting the reference test. Two studies stated that clinical information was available to the ultrasonographer,[36,77] three stated that it was not available[51,62–64] and it was not mentioned in the remaining studies.

The criteria used to interpret the ultrasound scans also tended to be a combination of those proposed in the early studies of Middleton and colleagues,[41,48] Mack and colleagues[46,47] and Crass and colleagues.[49,80] Almost all studies included non-visualisation of the RC as a criterion, 12[36,47,51,55,58,60,64,66,71–74] reported using loss of convexity of the RC as per Mack and colleagues[46] and 13 included focal discontinuity[41,45,53,54,60,62,63,69,70,73,75,77,78] and 18 focal thinning[41–44,54,57,59,61,64,67–70,72,73,75,76,78] as per Middleton and colleagues.[48] The usefulness of the presence of echogenic foci as an indicator of RCT as suggested by Crass and colleagues[40] is less accepted, but was included as a criterion in 10 studies.[40,42,44,54,56,67,69,70,75,76] Finally, the presence of an echogenic band as proposed in the early studies by Middleton and colleagues[48] was included in only four studies.[45,48,65,70] Two studies evaluated the use of the presence of fluid in the bursal system or subacromial–subdeltoid bursa.[50,52]

Studies rarely mentioned how they dealt with indeterminate results: five reported the number of indeterminate results that had been excluded,[40,41,43,70,72] three of which[40,41,43] also reported the reference test result for these patients so that it was possible to reanalyse their results, including the indeterminate results first as positive and then as negative. Another 16 studies[36,42–45,52,53,58,60–63,65,66,71,75] were judged not to have excluded any indeterminate results; however, it is possible in some cases and in those scored as 'unclear', that any indeterminate results were excluded before the sample was selected (e.g. in retrospective studies). Withdrawal bias did not appear to be present in any of the studies.

Results

The plot of sensitivity against 1 – specificity (false-positive rate) for each study and outcome are presented in *Figure 5*. These indicate a considerable range in sensitivity estimates for all outcomes, and a wide variation in specificity, particularly for the outcome of 'any tear'. Spearman correlations between sensitivity and specificity were not significant, and there was no obvious trend in the false-positive rates when ordered by the sensitivity rates (*Figure 6*). This indicated that it was reasonable to pursue a strategy of pooling the sensitivity and specificity rates, but the studies were highly heterogeneous for all outcomes. Subgroup analyses were conducted to explore reasons for differences between studies, but in general the heterogeneity remained (see *Tables 6–8*, discussed below).

Sensitivity and specificity

For any RCT, sensitivities ranged from 0.33 to 1.00 and specificities from 0.43 to 1.00. Sensitivity was lower in studies that were prospective in design (0.70, 95% CI: 0.64 to 0.75) compared with those that were retrospective or where the design was not stated (0.83, 95% CI: 0.81 to 0.85), although the difference was only just significant ($p = 0.05$) and significant heterogeneity remained in both subgroups. The only other significant differences between subgroups identified occurred when studies using ultrasound transducers of 7.5 MHz or less were compared to results in studies that used 10-MHz transducers on some or all patients (*Table 6*), although the number of studies in the latter group was very small. Statistical heterogeneity was removed when analyses were limited to low prevalence studies (pooled estimate 0.82, 95% CI: 0.73 to 0.88) and those with prevalences between 41 and 60% (pooled estimate 0.73, 95% CI: 0.69 to 0.77). Analysis of studies by year of publication as a possible proxy for changes in the test or increased experience with the test over time made very little difference to either sensitivity or specificity, despite an increase in mean prevalence with publication year.

For specificities, some significant differences according to age were identified. In those studies

TABLE 5 *Ultrasound: study methods and quality assessment results*

Study	Design[a]	Sample size	Mean age (years)	% males in sample	Frequency[b]	Criteria[c]	Ref[d]	Prevalence (%)[e]	Appropriate spectrum[f]	Eligibility criteria stated[f]	Appropriate ref. test[f]	Disease progression bias[f]	Partial verification[f]	Differential verification[f]	Incorporation bias[f]	Details of index test[f]	Details of ref. test[f]	Test review bias[f]	Diagnostic review bias[f]	Clinical info. available[f]	Indeterminate missing[f]	Withdrawal bias[f]
Ahovuo, 1989[73]	?	88	47	Nr	7.5	a, c, d	AG	32[F]	?	?	N	nr	N	N	N	Y	N	?	?	?	?	N
Arslan, 1999[50]	R	105	46	53	7.5	g	AG	49	?	Y	N	nr	?	N	N	Y	N	N	?	?	?	N
Brandt, 1989[70]	P3	72[g]	52	78	7.5/5	a,c,d,e,f	AG/S	48	N	?	?	N	Y	Y	N	Y	Y	N	N	?	Y	N
Brenneke, 1992[51]	?	120	Nr	Nr	5	a,b	AS	54	?	N	Y	nr	N	N	N	Y	N	N	?	N	?	N
Burk, 1989[71]	P1	23	Nr	Nr	5	a,b	AG	39	?	N	N	N	N	N	N	Y	Y	N	N	?	N	N
Chiou, 1999[76]	?	55	Nr	Nr	7/10	a,d,e,	AG/S	73	?	N	?	nr	Y	Y	N	Y	N	?	?	?	?	N
Crass, 1988[40]	R	124	Nr	Nr	10	e	S	53	?	N	Y	nr	Y	N	N	Y	N	N	?	?	Y	N
De Muynck, 1994[57]	?	38	53	66	7.5	a,d	AG	71	?	?	N	nr	N	Y	N	Y	N	?	?	?	?	N
Drakeford, 1990[68]	P2	50	51	69	7.5	a,d	AG	24	Y	Y	N	nr	?	N	N	Y	N	N	N	?	?	N
Farin, 1990[52]	?	102	45	58	7.5	g	S	21[I]	?	?	Y	nr	Y	N	N	Y	N	N	?	?	N	N
Gratz, 1998[43]	P2	17	50	52	nr	a,d	AG	35	?	?	N	nr	?	N	N	N	Y	?	?	?	Y	N
Hodler, 1988[79]	R	51	Nr	76	5/7.5	nr	S	69	?	?	Y	nr	Y	N	N	Y	N	N	?	?	?	N
Hodler, 1991[59]	P2	24	46	46	7.5	a,d	AG	63[F]	?	?	N	Y	?	Y	N	Y	Y	N	?	?	?	N
Kurol, 1991[44]	?	58	46	46	7.5	a,d,e	S	41	?	?	Y	nr	?	N	N	N	Y	N	?	?	N	N
Mack, 1988[47]	?	99	Nr	Nr	7.5/5	a,b	AG	49	?	Y	N	nr	N	Y	N	Y	N	N	?	?	?	N
Martin-Hervas, 2001[66]	P3	61	Nr	41	7.5	a,b	AS/S	56	?	Y	Y	?	Y	Y	N	Y	N	?	?	?	N	N
Middleton, 1986[41]	P3	106	47	71	10	a,c,d,f	AG	34	?	N	N	N	?	N	N	Y	Y	N	N	?	Y	N
Miller, 1989[72]	P1	56	55	54	5	a,b,d	AG	46	?	?	N	N	N	N	N	Y	N	Y	N	?	Y	N
Misamore, 1991[58]	P3	32	Nr	81	nr	a,b	AS/S	84	N	Y	Y	nr	Y	Y	N	Y	Y	N	N	?	N	N
Naredo, 1999[60]	P1	36	61	3	7.5	a,b,c	M	50	N	Y	N	N	N	N	N	Y	Y	N	Y	?	N	N
Nelson, 1991[64]	P1	19	42	76	5	b,d	AS/S	26	?	Y	Y	nr	N	Y	N	Y	N	N	Y	N	?	N
Olive, 1992[45]	P3	72	50	56	7.5	a,c,f	AG	40	?	?	N	N	Y	N	N	N	N	N	?	?	N	N
Paavolainen, 1994[61]	R	49	38	69	7.5	a,d	S	55	N	Y	Y	nr	?	N	N	Y	Y	N	N	?	N	N

continued

TABLE 5 *Ultrasound: study methods and quality assessment results (cont'd)*

Study	Design[a]	Sample size	Mean age (years)	% males in sample	Frequency[b]	Criteria[c]	Ref[d]	Prevalence (%)[e]	Appropriate spectrum[f]	Eligibility criteria stated[f]	Appropriate ref. test[f]	Disease progression bias[f]	Partial verification[f]	Differential verification[f]	Incorporation bias[f]	Details of index test[f]	Details of ref. test[f]	Test review bias[f]	Diagnostic review bias[f]	Clinical info. available[f]	Indeterminate missing[f]	Withdrawal bias[f]
Pattee, 1988[69]	?	52	47	81	7.5	a,c,d,e	AS	67	N	Y	Y	nr	?	N	N	Y	Y	?	?	?	?	N
Read, 1998[36]	?	42	44	Nr	7.5	a,b	AS/S	81[l]	?	Y	Y	Y	?	Y	N	Y	N	?	N	N	N	N
Roberts, 2001[77]	P3	24	49	67	7.5	a,c	AS	71	?	N	Y	nr	?	N	N	Y	Y	N	?	Y	?	N
Shipley, 1985[65]	?	12	65	Nr	5	a,f	AG	35	?	?	N	nr	?	N	N	Y	N	?	?	?	N	N
Soble, 1989[75]	?	75	Nr	Nr	7.5/5	a,c,d,e,	AG	46	?	?	N	N	?	N	N	Y	Y	?	?	?	N	N
Sonnabend, 1997[53]	?	110	49	71	7.5	a,c	AS/ S	63	?	Y	Y	Y	?	Y	N	Y	N	N	?	?	N	N
Swen, 1998[62]	P2	48	55	58	7.5/5	c	S	46	N	Y	Y	nr	?	N	N	Y	Y	?	N	N	N	N
Swen, 1999[63]	P2	21	54	57	7.5/5	c	AS	62	N	Y	Y	N	?	N	N	Y	Y	?	?	N	N	N
Takagishi, 1996[54]	?	122	51	63	7.5	c,d,e	S	48	?	?	Y	nr	?	Y	N	Y	Y	N	?	?	?	N
Teefey, 2000[55]	R	100	56	45	7.5/10	a,b	AS	80	N	Y	Y	Y	?	N	N	Y	N	N	Y	?	?	N
van Holsbeeck, 1995[78]	?	52	52	58	7.5	c,d	AS	73	?	Y	Y	Y	N	N	N	Y	Y	N	N	?	?	N
Van Moppes, 1995[67]	P3	41	Nr	41	7.5	a,d,e,	AS/S/AG	49	?	?	?	?	Y	Y	N	Y	N	N	Y	?	?	N
Vick, 1990[74]	?	81	Nr	nr	5	a,b	AG/S	30	?	Y	?	N	Y	Y	N	Y	Y	N	Y	?	?	N
Wallny, 2001[42]	P2	40	54	62	10	a,d,e	AS/S	58	?	Y	Y	nr	?	Y	N	Y	N	?	?	?	N	N
Wiener, 1993[56]	?	225	59	47	7.5	a,e	AS/S	70	?	?	Y	nr	?	Y	N	Y	Y	N	N	?	?	N

[a] Design: P1, prospective, unselected sample of patients suspected of rotator cuff disorder; P2, prospective, all patients referred for reference test; P3, prospective, selected patients underwent reference test; R, retrospective; ?, design not reported.
[b] Frequency of ultrasound transducer in MHz.
[c] Criteria for diagnosis: a, non-visualisation of RC; b, loss of convexity; c, focal discontinuity; d, focal thinning; e, echogenic foci; f, echogenic band; g, presence of bursal fluid.
[d] Reference test: AG, arthrography; AS, arthroscopy; S, surgery; M, MRI.
[e] Prevalence (%) of any tear, unless specified: F, full-thickness tear, I, impingement syndrome.
[f] Responses to quality assessment criteria: N, no; Y, yes; ?, cannot tell; nr, not reported.
[g] A further 26 underwent ultrasound but no reference test.

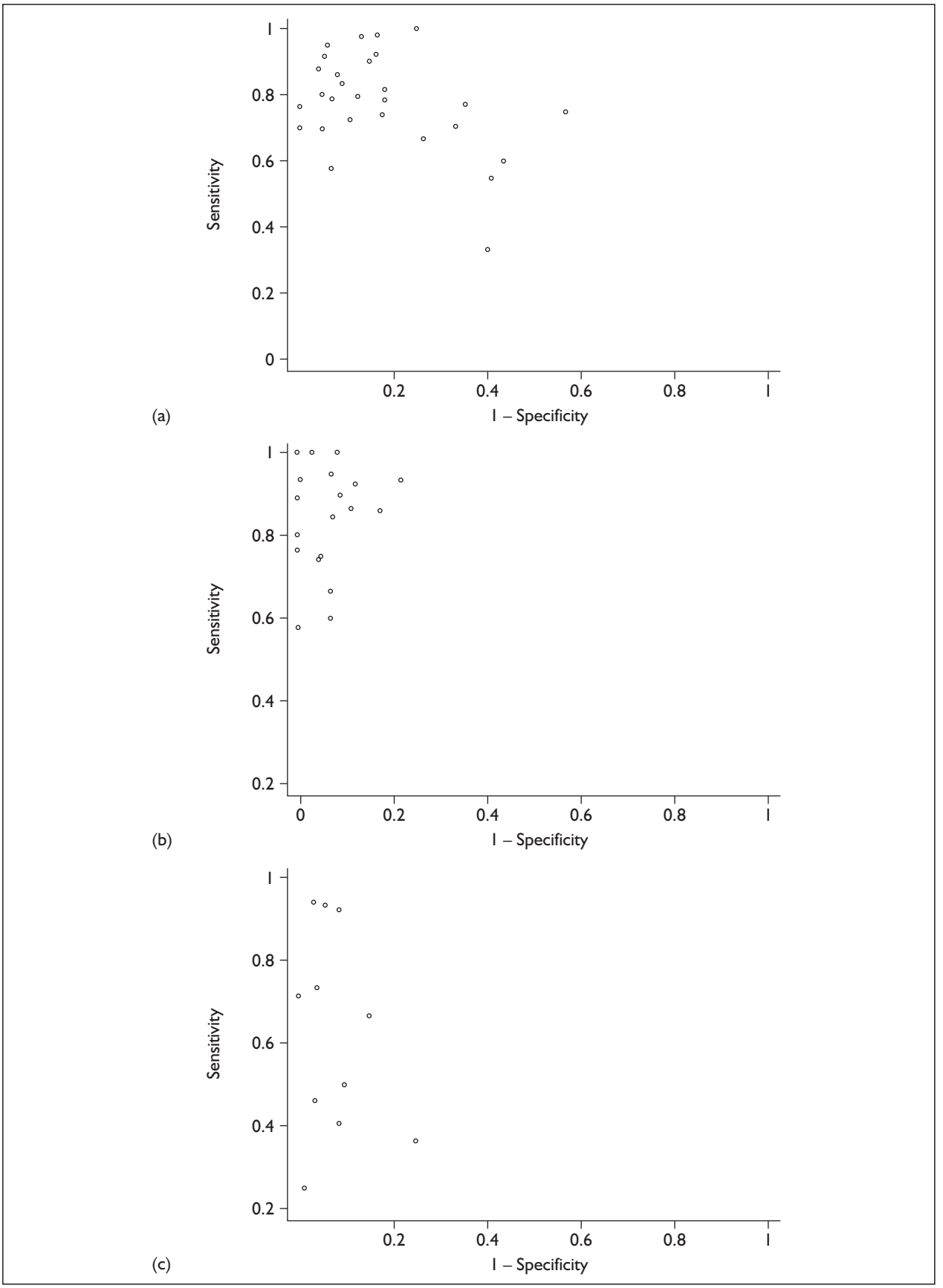

FIGURE 5 *Ultrasound studies: plot of sensitivity versus 1 – specificity for each outcome. (a) Detection of any tear* (n = *29): Spearman's rho = –0.2,* p = *0.29. (b) Detection of full-thickness tear* (n = *19): Spearman's rho = 0.1,* p = *0.69. (c) Detection of partial-thickness tear* (n = *11): Spearman's rho = –0.2,* p = *0.63*

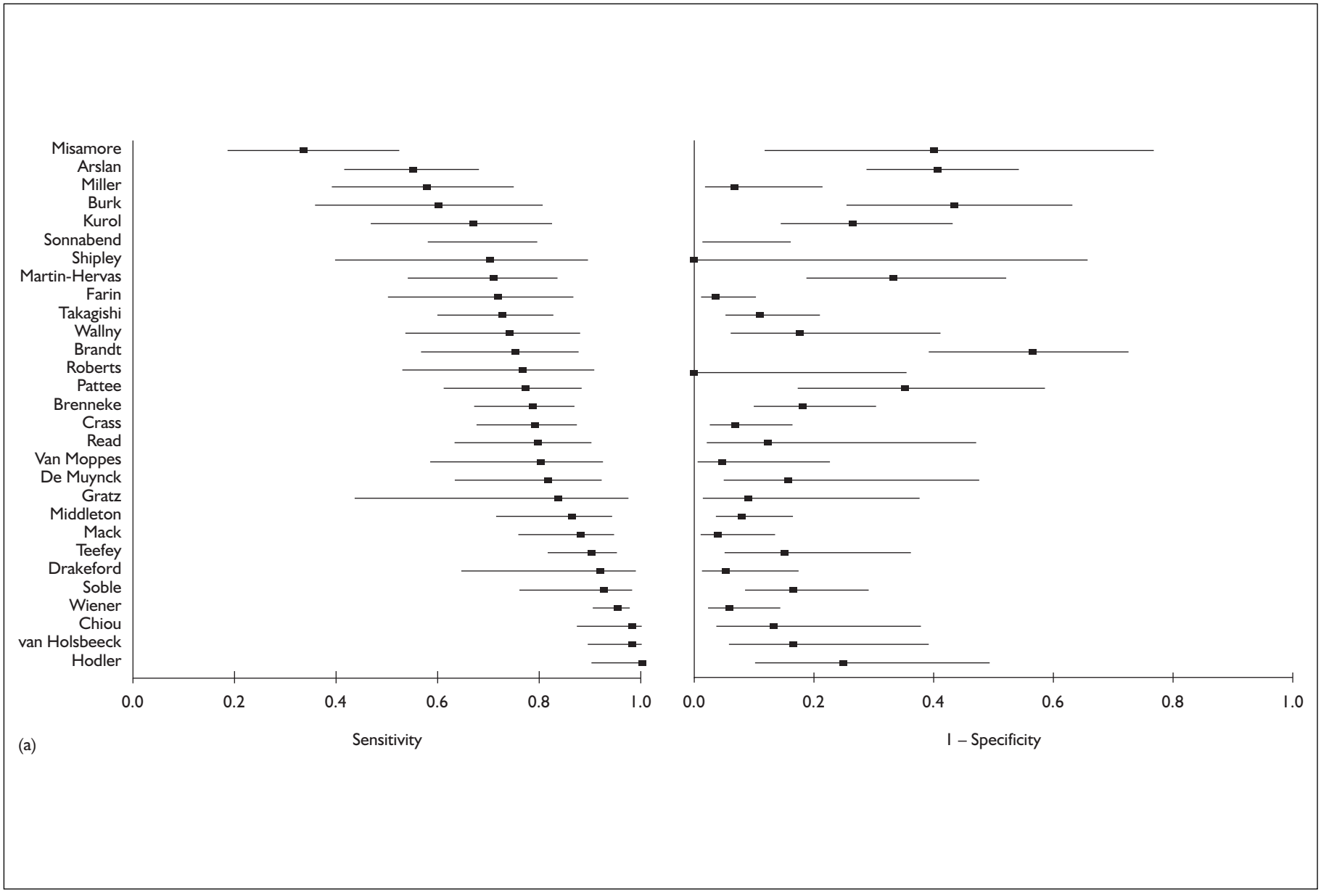

FIGURE 6 *Ultrasound: forest plots of sensitivity against false-positive rate (1 – specificity), per outcome. (a) Detection of any RCT* (n = *29) (cont'd)*

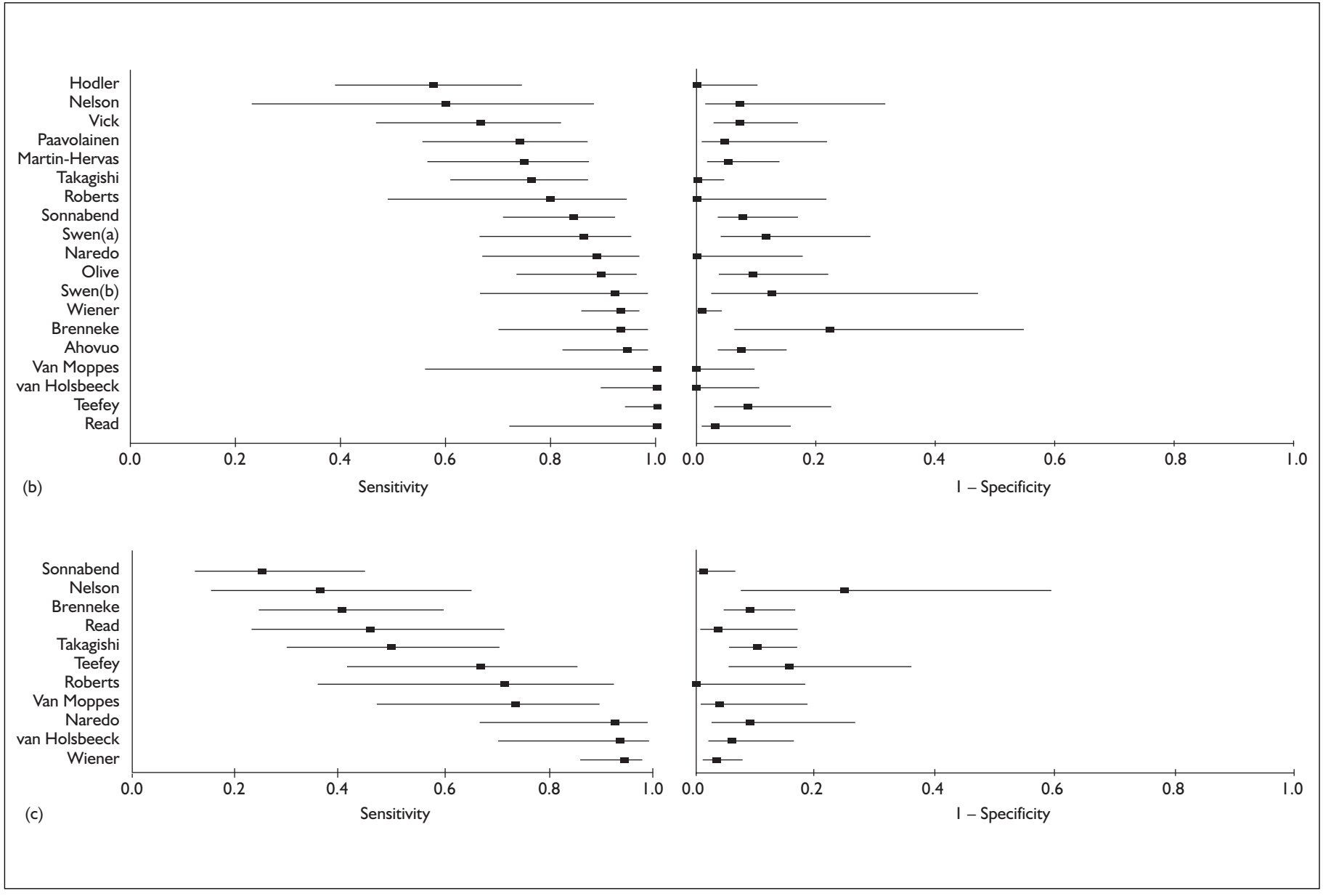

FIGURE 6 *(cont'd) (b) Detection of full-thickness RCT* (n = *19). (c) Detection of partial-thickness RTC* (n = *11)*

TABLE 6 *Ultrasound: summary sensitivity and specificity for detection of any tear*

Outcome subgroup	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
All studies	29	0.80 (0.78 to 0.83)*	0.85 (0.82 to 0.87)*
By age (years)			
<45	1	0.79 (0.63 to 0.90)	0.88 (0.53 to 0.98)
45–54	14	0.75 (0.71 to 0.79)*	0.83 (0.80 to 0.86)†*
55+	4	0.89 (0.85 to 0.92)*	0.92 (0.86 to 0.96)† ‡
Not reported	10	0.80 (0.76 to 0.84)*	0.84 (0.79 to 0.87)‡*
		NS differences	Test for diff.: $p < 0.01$†; $p = 0.02$‡
By prevalence (%)			
<40	6	0.82 (0.73 to 0.88)	0.89 (0.85 to 0.92)*
41–60	11	0.73 (0.69 to 0.77)	0.80 (0.76 to 0.84)*
>60	12	0.86 (0.82 to 0.88)*	0.87 (0.82 to 0.91)
		NS differences	NS differences
By publication year[a]			
Up to 1990	10	0.81 (0.76 to 0.85)*	0.83 (0.79 to 0.87)*
1990–1994	7	0.82 (0.78 to 0.86)*	0.89 (0.85 to 0.92)*
1995 onwards	12	0.79 (0.75 to 0.82)*	0.82 (0.77 to 0.86)*
		NS differences	NS differences
By frequency (MHz)			
⩽7.5	22	0.80 (0.77 to 0.83)†γ*	0.83 (0.80 to 0.86)*
7.5/10	2	0.92 (0.86 to 0.96)†‡	0.85 (0.71 to 0.94)
10 alone	3	0.80 (0.72 to 0.86)‡	0.91 (0.86 to 0.95)
7.5/10 or 10 alone	5	0.86 (0.81 to 0.90)γ*	0.90 (0.85 to 0.94)
NS	2	No pooled estimate	No pooled estimate
		Test for diff.: $p < 0.01$†‡; $p = 0.06$γ	NS differences
By reference test			
Arthrography	10	0.76 (0.70 to 0.80)*	0.84 (0.90 to 0.87)*
Arthroscopy	5	0.86 (0.81 to 0.90)*	0.81 (0.73 to 0.87)
Surgery	4	0.81 (0.74 to 0.86)*	0.89 (0.84 to 0.93)*
Arthroscopy/surgery	7	0.79 (0.74 to 0.82)*	0.88 (0.83 to 0.91)*
Other	3	No pooled estimate	No pooled estimate
		NS differences	NS differences
By design			
Prospective	11	0.70 (0.64 to 0.75)†*	0.81 (0.76 to 0.85)*
Retrospective (4) or not reported (14)	18	0.83 (0.81 to 0.85)†*	0.86 (0.83 to 0.88)*
		Test for diff.: $p = 0.05$†	NS differences
Two or more key biases[b]			
Absent	17	0.82 (0.79 to 0.84)*	0.84 (0.81 to 0.87)*
Present or not reported	12	0.78 (0.74 to 0.81)*	0.85 (0.81 to 0.89)
		NS differences	NS differences

[a] Mean prevalence of tear increased with each successive group of studies: 48, 53 and 61%, respectively.
[b] Key biases: partial verification bias; differential verification bias; test review bias; diagnostic review bias.
* Heterogeneity, $p < 0.05$ unless otherwise noted. Where the symbols †, ‡ and γ are paired or grouped they indicate that the specificity results are significantly different from each other.

with a higher mean age (aged 55 years or over), specificity was higher than where the mean age was 45–54 years: 0.92 (95% CI: 0.86 to 0.96) compared with 0.83 (95% CI: 0.80 to 0.86), test for difference $p < 0.01$. Heterogeneity in the former group was removed (although the number of studies was very small) but remained in the latter.

For full-thickness RCTs, overall pooled sensitivities and specificities were higher than for detection of any tear (*Table 7*), although significant heterogeneity remained: sensitivities ranged from 0.58 to 1.00 and specificities from 0.78 to 1.00. Sensitivity was significantly higher in high prevalence compared with low prevalence studies: 0.94 (95% CI: 0.90 to 0.96) compared with 0.75

TABLE 7 *Ultrasound: summary sensitivity and specificity for detection of full-thickness tears*

Outcome subgroup	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
All studies	19	0.87 (0.84 to 0.89)*	0.96 (0.94 to 0.97)*
By age (years)			
<45	3	0.79 (0.64 to 0.88)* ($p = 0.05$)	0.96 (0.91 to 0.97)
45–54	8	0.86 (0.81 to 0.90)*	0.95 (0.92 to 0.97)*
55+	4	0.94 (0.90 to 0.96)*	0.97 (0.93 to 0.98)*
Not reported	4	No pooled estimate	No pooled estimate
		NS differences	NS differences
By prevalence (%)			
<40	5	0.75 (0.66 to 0.82)†	0.96 (0.93 to 0.98)†
41–60	7	0.83 (0.76 to 0.87)*	0.95 (0.91 to 0.97)
>60	7	0.94 (0.90 to 0.96)†*	0.96 (0.93 to 0.97)†*
		Test for diff.: $p < 0.05$†	Test for diff.: $p < 0.05$†
By publication year[a]			
Up to 1994	8	0.86 (0.81 to 0.90)*	0.95 (0.92 to 0.97) ($p = 0.05$)
1995 onwards	11	0.88 (0.83 to 0.91)*	0.97 (0.94 to 0.98)*
		NS differences	NS differences
By frequency (MHz)			
5 alone	3	0.82 (0.71 to 0.89)*	0.93 (0.88 to 0.96)
5/7.5 MHz	2	0.89 (0.74 to 0.95)	0.88 (0.73 to 0.95)
⩾7.5 MHz	14	0.88 (0.84 to 0.90)	0.96 (0.95 to 0.98)
		NS differences	NS differences
By reference test			
Arthrography	3	0.85 (0.75 to 0.91)	0.92 (0.85 to 0.96)†
Arthroscopy	5	0.97 (0.93 to 0.98)†*	0.94 (0.90 to 0.97)
Surgery	2	0.80 (0.66 to 0.89)†‡	0.92 (0.80 to 0.97)
Arthroscopy/surgery	6	0.84 (0.78 to 0.88)‡*	0.98 (0.96 to 0.99)†*
Other	3	0.78 (0.65 to 0.88)	0.96 (0.91 to 0.98)
		Test for diff.: $p < 0.05$†‡	Test for diff.: $p < 0.05$†
By design			
Prospective	9	0.83 (0.75 to 0.88)*	0.94 (0.91 to 0.97)*
Retrospective (2) or not reported (8)	10	0.88 (0.84 to 0.91)*	0.96 (0.94 to 0.97)*
		NS differences	NS differences
Two or more key biases[b]			
Absent	11	0.91 (0.87 to 0.93)*	0.95 (0.93 to 0.97)*
Present or not reported	8	0.79 (0.72 to 0.84)*	0.96 (0.93 to 0.98)*
		NS differences	NS differences

[a] Mean prevalence of tear increased slightly with publication year: 38 and 43%, respectively.
[b] Key biases: partial verification bias; differential verification bias; test review bias; diagnostic review bias.
* Heterogeneity, $p < 0.05$ unless otherwise noted. Where the symbols †, ‡ and γ are paired or grouped they indicate that the specificity results are significantly different from each other.

(95% CI: 0.66 to 0.82). Sensitivity also appeared to increase with increasing mean age, although differences between subgroups were not found to be statistically significant. Sensitivity was lower in those studies using arthrography as the reference test than those using arthroscopy, but specificity estimates were less affected. Insufficient numbers of studies using transducers of >7.5 MHz were available to perform any meaningful subgroup analysis on that basis, but there was an indication that accuracy improved when using 7.5 MHz as opposed to 5 MHz. The subgroup analyses by year did not demonstrate any real change in accuracy over time.

For specificities, some statistically significant differences according to frequency of ultrasound scanner and reference test used were found, but again the number of studies in some subgroups was small, limiting the power of these comparisons.

TABLE 8 *Ultrasound: summary sensitivity and specificity for detection of partial-thickness tears*

Outcome subgroup	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
All studies	11	0.67 (0.61 to 0.73)*	0.94 (0.92 to 0.96)*
By age (years)			
<45	2	0.42 (0.24 to 0.61)†*	0.92 (0.79 to 0.97)
45–54	4	0.53 (0.41 to 0.65)*	0.95 (0.91 to 0.97)*
55+	3	0.89 (0.82 to 0.94)†*	0.95 (0.91 to 0.97)
Not reported	2	No pooled estimate	No pooled estimate
		Test for diff.: $p < 0.01$†	NS differences
By prevalence (%)			
<40	3	0.45 (0.32 to 0.60)†‡	0.91 (0.85 to 0.94)
41–60	3	0.62 (0.49 to 0.74)†*	0.92 (0.87 to 0.96)
>60	5	0.77 (0.69 to 0.83)‡*	0.96 (0.94 to 0.98)
		Test for diff.: $p = 0.06$†, $p = 0.02$‡	
By publication year[a]			
Up to 1994	3	0.75 (0.65 to 0.82)*	0.94 (0.91 to 0.96)*
1995 onwards	8	0.61 (0.52 to 0.69)*	0.94 (0.91 to 0.96)
		NS differences	NS differences
By frequency (MHz)			
5	2	0.39 (0.26 to 0.55)†	0.90 (0.83 to 0.95)
7.5 or 7.5/10	9	0.73 (0.66 to 0.78)†*	0.95 (0.93 to 0.96)
		Test for diff.: $p < 0.01$†	NS differences ($p = 0.09$)
By reference test			
Arthroscopy	4	0.63 (0.50 to 0.73)†*	0.92 (0.88 to 0.95)
Arthroscopy/surgery	6	0.67 (0.59 to 0.74)‡*	0.95 (0.93 to 0.97)*
Other (MRI)	1	0.92 (0.67 to 0.99)†‡	0.91 (0.73 to 0.98)
		Test for diff.: $p < 0.05$†‡	NS differences
By design			
Prospective	4	0.70 (0.55 to 0.81)*	0.94 (0.92 to 0.96)
Retrospective (1) or not reported (6)	7	0.66 (0.59 to 0.73)*	0.93 (0.85 to 0.97)*
		NS differences	NS differences
Two or more key biases[b]			
Absent	7	0.77 (0.70 to 0.83)†*	0.94 (0.91 to 0.96)
Present or not reported	4	0.46 (0.35 to 0.57)†*	0.95 (0.91 to 0.97)* ($p = 0.05$)
		Test for diff.: $p = 0.05$†	NS differences

[a] Mean prevalence of tear decreased slightly with increasing publication year: 29 and 26%, respectively.
[b] Key biases: partial verification bias; differential verification bias; test review bias; diagnostic review bias.
* Heterogeneity, $p < 0.05$ unless otherwise noted. Where the symbols † and ‡ are paired or grouped they indicate that the specificity results are significantly different from each other.

For detection of partial-thickness RCTs (*Table 8*), the pooled sensitivity estimate was low (0.67, 95% CI: 0.61 to 0.73) although specificity remained high (0.94, 95% CI: 0.92 to 0.96), and studies were again very heterogeneous, especially in terms of sensitivity (see *Figure 6*), which ranged from 0.25 to 0.94. Statistically, several possible reasons for the differences in sensitivity estimates between studies were identified, including age, prevalence, scanner frequency and absence of key biases, giving limited power to these comparisons. No significant differences in terms of specificity estimates were identified.

LRs

For the detection of a full-thickness RCT, a positive ultrasound result increases the odds of RCT being present by ~13-fold (positive LR, 13.16; 95% CI: 9.13 to 18.95; *Table 9*). This is just above the level usually required for a diagnostic test to provide convincing diagnostic evidence. The LRs from the 19 studies contributing to this estimate were statistically homogeneous ($p = 0.2$). *Figure 7* demonstrates that the LRs from each of the studies were significant (CIs did not cross one) and all of the studies, except that of Wiener and Seitz,[56] included the overall pooled estimate in their CIs.

TABLE 9 *Ultrasound: summary likelihood ratios*

	No. of studies	Positive LR (95% CI)	Negative LR (95% CI)
Any tear	29	5.09 (3.44 to 7.53)*	0.27 (0.19 to 0.34)*
Full tear	19	13.16 (9.13 to 18.95)	0.16 (0.11 to 0.24)*
Partial tear	11	8.90 (4.88 to 16.22)*	0.36 (0.22 to 0.61)*

* Heterogeneity, $p < 0.01$.

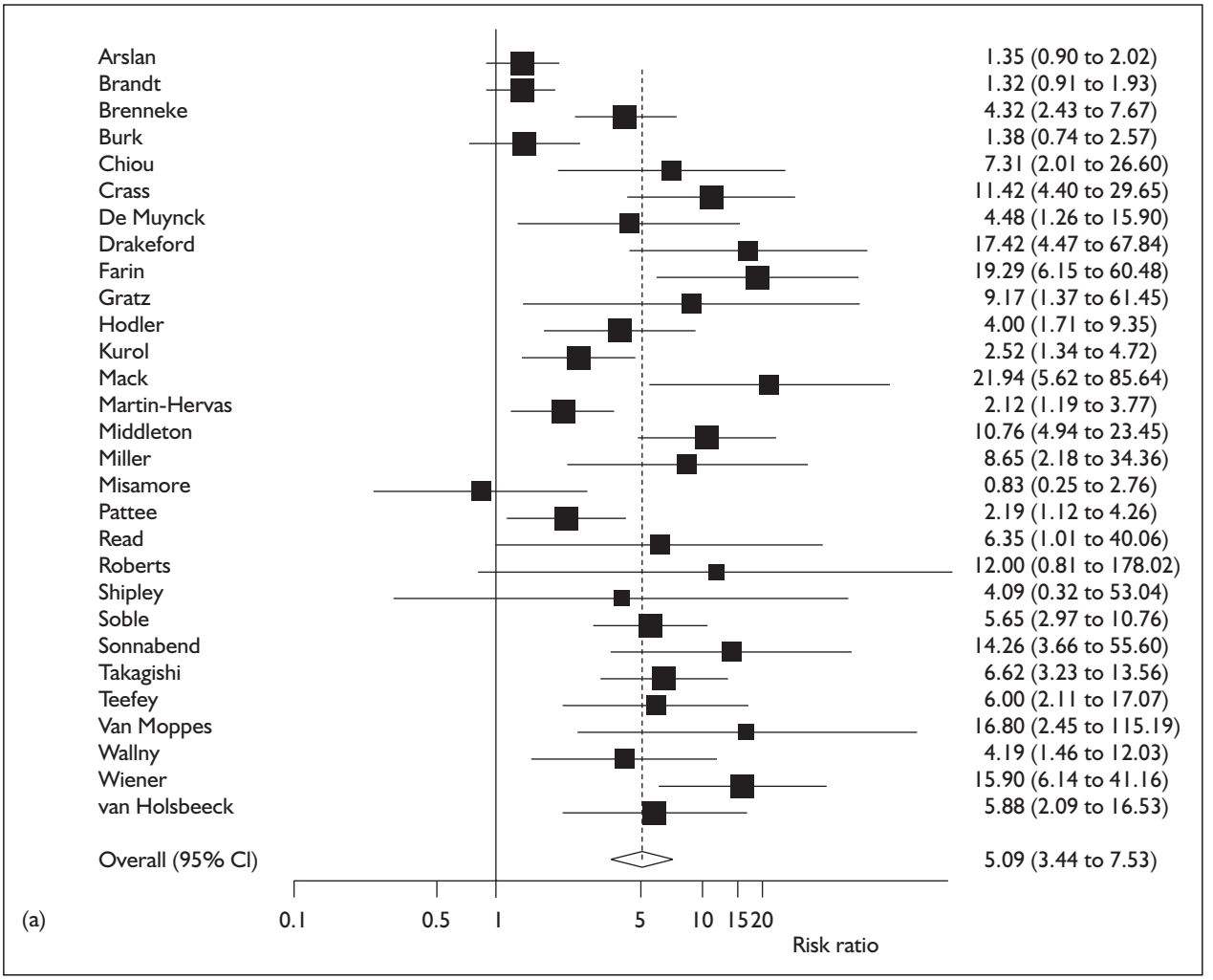

FIGURE 7 *Ultrasound: forest plots of LRs, per outcome. (a) Any tear: positive LRs (cont'd)*

For the detection of any tear or partial-thickness tears, the odds of RCT being present when a positive test result was shown increased by around 5- and 9-fold, respectively (*Table 9*). For these comparisons, however, the studies were much more heterogeneous ($p < 0.01$ for both). The forest plots in *Figure 7* show that the positive LRs were particularly spread out for detection of any tear, several were not significant and nine did not include the overall summary estimate in their CIs.

In terms of negative LRs (the odds of a negative test result in a person who does have the target disorder), ultrasound again performed the best when focused on detection of full-thickness RCTs, although the studies were statistically heterogeneous (negative LR, 0.16; 95% CI: 0.11 to 0.24). Nevertheless, it still does not perform at a level that provides convincing diagnostic evidence and will only be useful when applied in high prevalence settings, that is, in hospital clinics and perhaps only those with the more difficult cases.

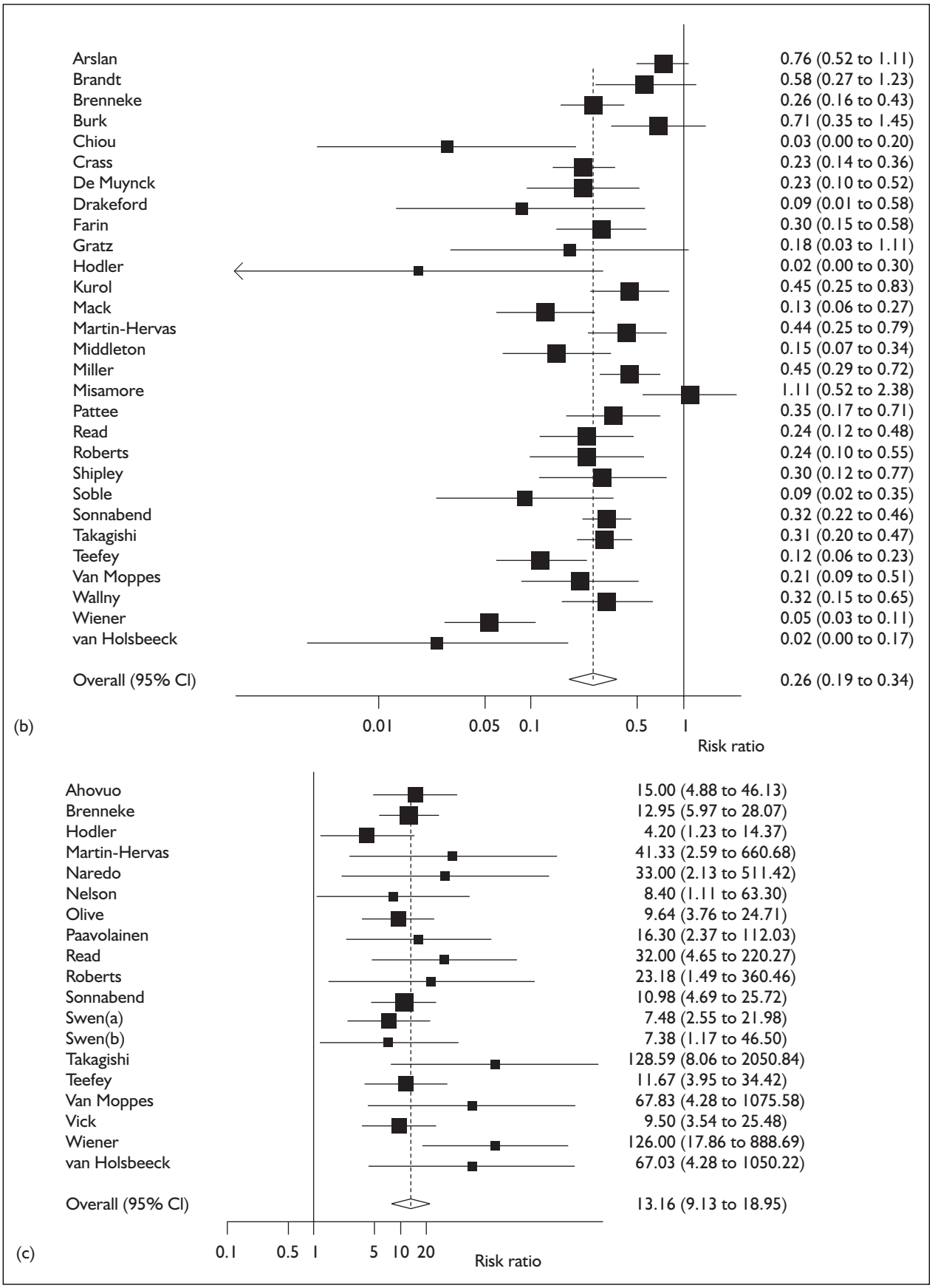

FIGURE 7 *(cont'd) (b) Any tear: negative LRs. (c) Full-thickness tear: positive LRs (cont'd overleaf)*

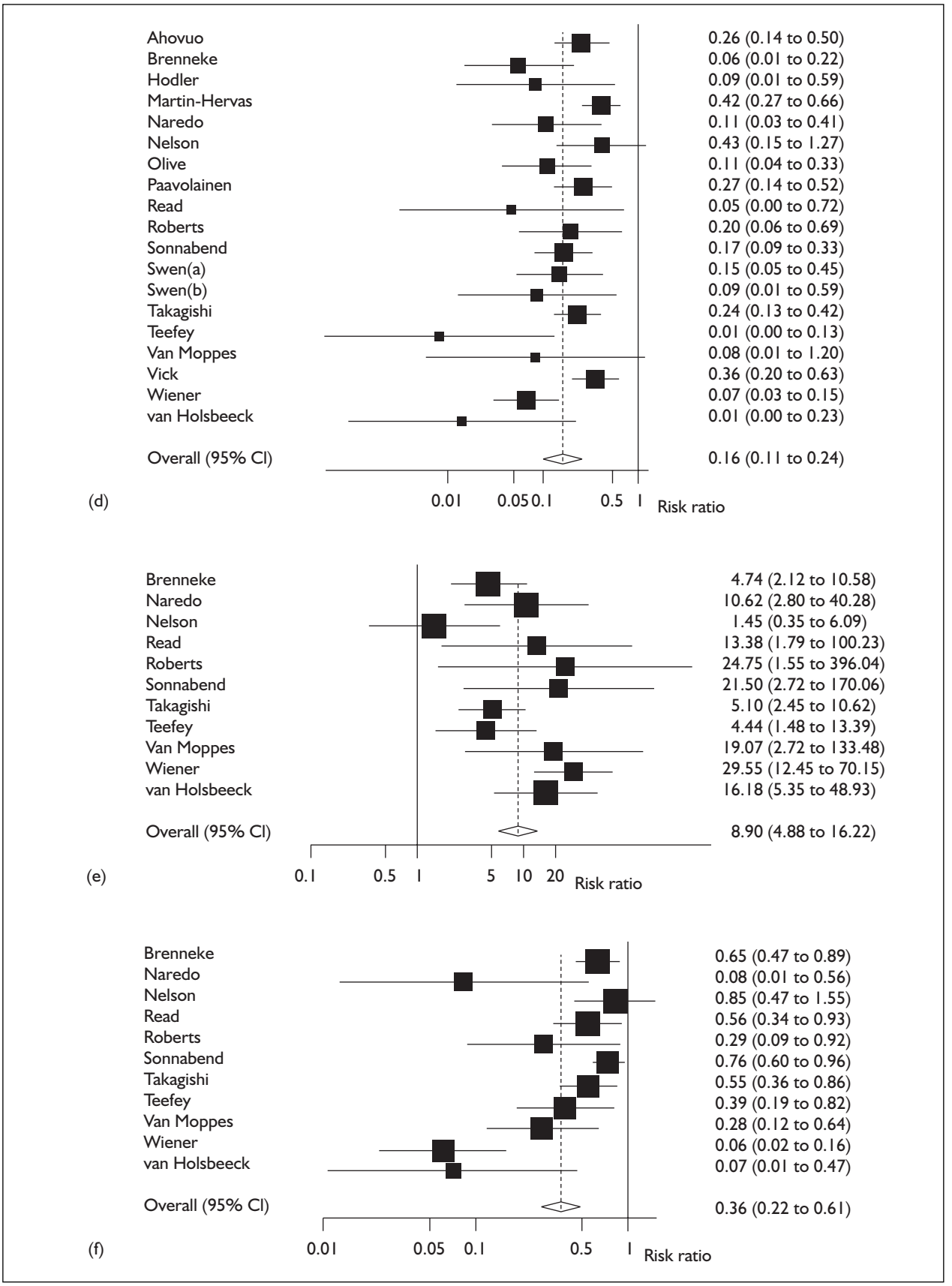

FIGURE 7 *(cont'd) (d) Full-thickness tear: negative LRs. (e) Partial-thickness tears: positive LRs. (f) Partial-thickness tears: negative LRs*

Comparisons within individual studies

Three studies (by Crass and colleagues,[40] Gratz and colleagues[43] and Middleton and colleagues[41]) reported the number of indeterminate results that occurred along with the diagnosis from the reference test. All excluded the indeterminate results from their estimations of sensitivity and specificity, but the data allowed the results to be recalculated including the indeterminates as either positive or negative. The data presented for these studies in Appendix 7 show that such a policy artificially inflates sensitivity and/or specificity.

A further retrospective study[70] presented results first using the original, or prospective, interpretation of the ultrasound scan and then using the new interpretation from retrospective review of the ultrasound scans. Sensitivity was slightly higher and specificity much lower when the prospective interpretation was used.

Two studies evaluated the use of different criteria for the interpretation of the ultrasound scans. Arslan and colleagues[50] evaluated the presence of biceps effusion and/or bursal fluid as an indicator of RCT; neither was found to have high diagnostic power. Kurol and colleagues[44] examined whether changes in echogenicity alone or cuff thinning alone were sufficient indicators of tear. Both were found to have low sensitivity and slightly higher specificity (82 or 88%) when used alone, but even when either or both were considered as indicators of tear, ultrasound was found to have low accuracy.

One recent study[42] compared accuracy using two-dimensional (2D) versus three-dimensional (3D) ultrasonography. The advantage of the 3D method is said to be the dynamic multiplanar imaging facility that allows the observation of three perpendicular planes. However, 2D imaging needs to be performed first in order to define the region of interest. The authors found the sensitivity of 3D imaging to be more sensitive but similarly specific to 2D imaging in 40 patients. Positive LRs were low for both tests, but the negative LR was better for 3D than 2D imaging: 0.1 (95% CI: 0.0 to 0.1) versus 0.3 (95% CI: 0.2 to 0.7). These results were not independent as the tests were conducted and interpreted sequentially by the same ultrasonographer.

Only one study[63] reported results for more than one reader. Both interpreters were said to be experienced, but one had a much higher sensitivity than the other (92 versus 69%), although specificities were similar (Appendix 7).

Discussion

The results indicate a wide variation in accuracy across individual studies. Sensitivity and specificity were highly heterogeneous for each of the three outcomes investigated, although perhaps less so for detection of full-thickness RCTs. The only homogeneous finding was the pooled positive LR estimate for detection of full-thickness tears. It is likely that this is because the diagnostic criteria for full-thickness tears are more much more clearly defined than for partial tears. As discussed previously (see the section 'Criteria for interpretation of ultrasound scans', p. 6), sonographic criteria for partial tears overlap with those for full-thickness tears. The variability in, and poor reporting of, diagnostic criteria used between studies made it impossible to investigate further those criteria that may be most accurate at identifying tears on ultrasound. It is also very difficult to draw any conclusions regarding the effect of increasing frequency of ultrasound transducers. Although increasing frequency (especially with machines of 10 MHz or more) increases image resolution, with only three studies identified that employed only 10 MHz transducers, there is insufficient evidence as to how this might translate into improved accuracy. We subdivided studies according to year of publication, as a proxy for improvements in the technique or operator experience over time, but did not identify any obvious trends in either sensitivity or specificity.

Nevertheless, our results do demonstrate that a positive ultrasound finding of a full-thickness tear in these studies may provide convincing evidence that such a tear is in fact present, increasing the odds by around 13-fold. Using Bayes theorem, the pooled LR can be applied to the pre-test probability of disease in these studies (average prevalence across studies) of 53%, increasing it to over 90% (*Box 1*). It would therefore appear that when used in similar settings to the studies identified, patients with positive ultrasound findings could be considered as potential candidates for surgical treatment without further investigation, if their symptoms are sufficiently severe. It remains to be determined whether or not ultrasound can provide such conclusive evidence for the value of a negative ultrasound finding in ruling out the presence of a full-thickness tear. Studies to date have been too heterogeneous to allow any strong conclusion to be drawn.

It should also be borne in mind that all of the studies were conducted in samples with a relatively high prevalence of RCT and are therefore likely to

BOX 1 *Ultrasound: impact of a positive test result on the probability of full-thickness tear*

Summary positive LR for full-thickness tear (across 19 studies):		13.16
Pre-test probability of FT tear in these studies (average prevalence):		0.53
Pre-test odds of FT tear	= 0.53/(1 – 0.53)	= 1.13
Post-test odds of FT tear	= 1.13 × 13.16	= 14.84
Post-test probability of FT tear	= 14.84/(1 + 14.84)	= 0.94

A positive ultrasound finding of FT tear increases the probability of such a tear being present from around 50% to over 90%

have included a high proportion of more severe cases. A large number of studies included only those who had experienced shoulder pain for a reasonable period of time and those who had already failed conservative treatment or were recognised 'problem' patients. Even in those studies classified as 'low prevalence', between 24 and 32% of patients were found to have full-thickness tears. Accuracy is likely to be lower in lower prevalence settings, as was suggested by the subgroup analyses.

There was also a suggestion that sensitivity was higher in studies where the mean age of participants was higher. It may be that in these studies, patients with alternative causes of shoulder pain may have been more easily excluded from the study sample, for example, those with arthritis. Alternatively, ultrasound may be better at picking up RC disorders that result from classical outlet impingement, more common in older patients. There could also be some confounding between age and prevalence: full-thickness tears in particular are more common in the elderly, hence studies of an older age group of patients would tend to have a higher prevalence of tears, and therefore tests might demonstrate better accuracy.

Furthermore, all of the comparisons are confounded by small sample sizes and the poor quality and reporting of included studies, especially regarding key biases such as verification bias and blinding. Interestingly, sensitivity improved where two or more of these biases were reported to be absent compared with those where they were present or not reported, although both subgroups of studies remained heterogeneous. One might usually expect accuracy to be lower in better designed studies. Accuracy also tends to be underestimated when an imperfect reference test is used;[81] in this case, sensitivity was markedly lower in those studies using arthrography as a reference test, especially for detection of any tear, although specificity was less affected. Of further note is the lack of reporting of indeterminate or unclear results. A certain proportion of such results will occur with any test, especially where there is an obvious subjective component, as with imaging tests. As discussed above, the exclusion of indeterminate results can markedly increase sensitivity and/or specificity.

In conclusion, then, ultrasound may be able to rule in full-thickness tears and therefore rapidly identify those requiring surgical treatment, but this has as yet only been demonstrated in studies conducted in specialist settings where the prevalence of disease is high. Its diagnostic capabilities in more routine settings and its ability to rule out disease effectively when a negative result is shown is questionable and remains to be determined. There is as yet no conclusive evidence for its ability to rule in or out the presence of partial-thickness tears, but it seems likely that its ability to do either is low.

MRI

Description and quality of included studies

A total of 29 cohort studies investigating the accuracy of MRI for the diagnosis of RC disorders were identified. Summary details of the methods and results are provided in *Tables 10–13*. Full study details and results are given in Appendices 9–11.

Interventions

Two studies[38,82] did not provide any details of the MRI used. Of the remainder, the majority (22) evaluated conventional MRI pulse sequences against a reference test, two[83,84] evaluated fat-suppressed MRI against a reference test and one[85] used a mixture of conventional and fat-suppressed MRI. Two further studies compared conventional MRI with or without fat suppression against a reference test.[86,87]

TABLE 10 *MRI: study methods and quality assessment results*

Study	Design[a]	Sample size	Mean age (years)	% male participants	MRI details[b]	Ref[c]	Prevalence[d]	Appropriate spectrum[e]	Eligibility criteria stated[e]	Appropriate ref. test[e]	Disease progression bias[e]	Partial verification[e]	Differential verification[e]	Incorporation bias[e]	Details of index test[e]	Details of ref. test[e]	Test review bias[e]	Diagnostic review bias[e]	Clinical info. available[e]	Indeterminate missing[e]	Withdrawal bias[e]
Balich, 1997[83]	R	222	45	58	Fat; OC, OS, AX	AS	32	?	Y	Y	Y	N	N	N	Y	Y	N	?	N	?	N
Birtane, 2001[88]	P1	125	52	40	Con; OC	SIT	70[I]	?	Y	?	?	N	N	N	Y	Y	?	?	?	?	N
Blanchard, 1999[98]	P3	38	50	55	Con; OC, OS, AX	AS/S	29[F]	?	?	Y	?	Y	Y	N	Y	N	N	?	?	?	Y
Burk, 1989[71]	P1	38	nr	nr	Con; OC, OS, AX	AG	58	?	N	N	N	N	N	N	Y	Y	N	N	?	?	N
Evancho, 1988[89]	?	28	nr	nr	Con; OC	AS/AG	42	?	N	?	?	N	Y	N	Y	Y	N	?	?	N	Y
Hodler, 1991[59]	P2	24	58	43	Con; OC	AG	63[F]	?	N	N	?	N	N	N	Y	Y	?	?	?	?	N
Hodler, 1992[90]	?	36	42	67	Con; OC	AS	47	?	N	Y	?	Y	N	N	Y	Y	N	?	N	N	N
Iannotti, 1991[17]	R	88	40	Nr	Con; OC, OS, AX	AS/S	35[I]	N	Y	N	?	Y	Y	N	Y	Y	?	?	N	?	Y
Jaovisidha, 1999[91]	P1	8	nr	nr	Con; OC	AS	25	?	N	Y	?	Y	N	N	Y	Y	N	?	N	?	N
Kieft, 19881[13]	P2	10	nr	nr	Con; OC, AX	AG	30	?	N	N	?	N	N	N	Y	Y	N	?	Y	?	N
Kneeland, 1987[95]	R	26	nr	60	Con; OC, AX	AG/S	85	?	N	?	N	N	Y	N	Y	N	Y	Y	?	?	N
Martin-Hervas, 2001[66]	P3	61	nr	41	Con; OC, AX	AS/S	56	?	N	Y	?	N	Y	N	Y	N	?	?	?	?	N
Morrison, 1990[94]	P1	100	nr	nr	Con; OS	AG	55	N	N	N	?	N	N	N	Y	Y	N	N	N	?	N
Nelson, 1991[64]	P1	21	42	76	Con; OC, AX	AS/S	90	N	N	Y	?	N	Y	N	Y	N	N	?	N	?	N
Patten, 1994[92]	R	50	nr	nr	Con; OC vs OC, OS	AS/AT/AG	40	?	N	?	?	N	Y	N	Y	N	N	?	N	N	N
Quinn, 1995[84]	R	100	47	54	Fat; OC, OS, AX	AS	31	?	N	Y	?	Y	Y	N	Y	Y	N	Y	?	?	N
Reinus, 1995[86]	?	49	37	67	Con vs Fat; OC, AX	AS	20[F]	N	N	Y	Y	Y	Y	N	Y	Y	N	?	N	?	N
Robertson, 1995[85]	R	82	45	55	Mix; OC, OS	S	31[F]	N	N	Y	Y	Y	N	N	Y	Y	N	?	N	?	Y
Sahin-Akyar, 1998[87]	R	39	54	44	Con vs Fat; OC	AS/AT	62	?	N	Y	Y	Y	Y	N	Y	Y	N	Y	N	?	Y
Swen, 1999[63]	P2	21	54	57	Con; OC	AS	62[F]	?	Y	Y	N	N	N	N	Y	Y	N	?	N	?	N

continued

TABLE 10 *MRI: study methods and quality assessment results (cont'd)*

Study	Design[a]	Sample size	Mean age (years)	% male participants	MRI details[b]	Ref[c]	Prevalence[d]	Appropriate spectrum[e]	Eligibility criteria stated[e]	Appropriate ref. test[e]	Disease progression bias[e]	Partial verification[e]	Differential verification[e]	Incorporation bias[e]	Details of index test[e]	Details of ref. test[e]	Test review bias[e]	Diagnostic review bias[e]	Clinical info. available[e]	Indeterminate missing[e]	Withdrawal bias[e]
Torstensen, 1999[82]	R	57	41	58	nr; nr	AS	42	N	N	Y	?	N	N	N	N	Y	?	?	Y	?	N
Traughber, 1992[114]	R	28	nr	nr	Con; OC, OS, AX	AS/S	50	?	N	Y	?	N	Y	N	Y	Y	N	?	?	?	N
Tuite, 1994[99,115]	R	100	42	71	Con; OC, OS	AS	56	?	N	Y	?	N	N	N	Y	Y	N	Y	?	?	N
Tuite, 2001[93]	R	75	40	65	Con; OC, OS vs OC, AS	AS/BS	65	?	N	Y	Y	Y	N	N	Y	N	N	?	N	N	N
Wang, 1994[96]	R	40	49	67	Con; OC, OS, AX	AS/S	55	?	N	Y	N	N	Y	N	Y	N	N	?	N	?	N
Wnorowski, 1997[37]	R	39	30	nr	Con; OC, OS, AX	AS	36	N	N	Y	Y	N	N	N	Y	N	?	?	Y	?	N
Wolf, 2001[38]	?	71	51	61	nr; nr	AS	46	?	N	Y	?	N	N	N	Y	N	N	?	?	?	N
Yagci, 2001[97]	P2	24	52	29	Con; OC, OS, AX	AS/S	63	?	N	Y	N	N	Y	N	Y	N	N	?	N	?	N
Zlatkin, 1989[16]	?	32	nr	nr	Con; OC, OS, AX	S	69	N	Y	Y	?	Y	Y	N	Y	Y	N	?	N	?	N

[a] Design: P1, prospective, unselected sample of patients suspected of RC disorder; P2, prospective, all patients referred for reference test; P3, prospective, selected patients underwent reference test; R, retrospective; ?, design not reported.
[b] MRI details: Con, conventional MRI; Fat, fat-suppressed MRI; OC, oblique coronal view; OS, oblique sagittal view; AX, axial view; AS, angled sagittal view.
[c] Reference test: AG, arthrography; AS, arthroscopy; AT, arthrotomy; S, Surgery; SIT, subacromial injection test.
[d] Prevalence (%) of any tear, unless specified: F, full-thickness tear; I, impingement syndrome.
[e] Responses to quality assessment criteria: N, no; Y, yes; ?, cannot tell.

TABLE 11 *MRI: summary results for detection of any tear*

Outcome subgroup	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
All studies	20	0.83 (0.79 to 0.86)*	0.86 (0.83 to 0.88)*
By type of MRI[a]			
Conventional	18	0.85 (0.81 to 0.88)*	0.80 (0.75 to 0.84)*
Fat-suppressed	3	0.78 (0.69 to 0.84)*	0.93 (0.89 to 0.96)*
		NS differences	NS differences
By publication year[b]			
Up to 1990	5	0.90 (0.82 to 0.95) ($p = 0.05$)	0.84 (0.72 to 0.91)
1990–94	7	0.88 (0.83 to 0.92)*	0.90 (0.85 to 0.93)*
1995+	8	0.76 (0.70 to 0.81)*	0.82 (0.78 to 0.86)*
		NS differences	NS differences
By age (years)			
<45	5	0.80 (0.73 to 0.85)†‡*	0.70 (0.62 to 0.77)* ($p = 0.05$)
45–54	5	0.74 (0.66 to 0.80)†* ($p = 0.05$)	0.93 (0.89 to 0.95)*
Not reported	10	0.92 (0.88 to 0.95)‡*	0.88 (0.83 to 0.92)
		Test for diff.: $p = 0.06$†; ($p = 0.01$)‡	NS differences
By prevalence (%)			
<40	5	0.74 (0.66 to 0.93)	0.93 (0.89 to 0.95)†*
41–60	10	0.90 (0.85 to 0.93)*	0.81 (0.76 to 0.86)†‡*
>60	5	0.77 (0.69 to 0.83)* ($p = 0.03$)	0.75 (0.63 to 0.84)‡
		NS differences	Test for diff.: $p = 0.03$†; $p = 0.05$‡
By reference test			
Arthrography	3	1.00 (0.44 to 1.00)	0.92 (0.82 to 0.96)
Arthroscopy	8	0.78 (0.72 to 0.82)*	0.85 (0.82 to 0.89)*
Surgery (1) or arthroscopy/surgery	4	0.85 (0.76 to 0.91)*	0.75 (0.63 to 0.84)†
Other	5	0.81 (0.71 to 0.87)	0.90 (0.82 to 0.95)†
			Test for diff.: $p = 0.03$
By views used[c]			
OC, Sag, AX	8	0.78 (0.72 to 0.83)†‡*	0.91 (0.87 to 0.93)†*
OC, Sg	3	0.86 (0.78 to 0.91)†*	0.82 (0.73 to 0.88)‡
OC, AX	3	0.92 (0.82 to 0.96)δ	0.74 (0.58 to 0.85)‡δγ
OC	5	0.68 (0.57 to 0.78)δγ* ($p = 0.04$)	0.84 (0.75 to 0.90)δ
Other	2	0.93 (0.87 to 0.97)‡γ*	0.94 (0.86 to 0.98)†γ
Not reported	1	0.96 (0.80 to 0.99)	0.48 (0.33 to 0.65)
		Test for diff.: $p < 0.01$†; $p = 0.06$‡; $p = 0.02$δ; $p = 0.05$γ	Test for diff.: $p = 0.01$†; $p = 0.04$‡; $p < 0.01$δ; $p < 0.01$γ
By design			
Prospective	6	0.92 (0.87 to 0.96)*	0.85, 0.77 to 0.90)*
Retrospective or not reported (3)	14	0.80 (0.75 to 0.83)*	0.86 (0.82 to 0.89)*
		NS differences	NS differences
Key biases[d]			
1 reported absent	5	0.86 (0.79 to 0.91)	0.87 (0.80 to 0.92)*
2 reported absent	10	0.76 (0.69 to 0.81)†*	0.78 (0.72 to 0.83)*
3 or 4 reported absent	5	0.87 (0.82 to 0.91)†*	0.90 (0.86 to 0.93)*
		Test for difference: $p = 0.03$	NS differences

[a] Total number of studies adds to 21 because one study (Sahin-Akyar and colleagues[87]) compared conventional and fat-suppressed MR.
[b] Mean prevalence of tear decreases with each successive group of studies: 60, 47, 44%, respectively.
[c] Total number of studies adds to 22 because two studies compared two sets of views (Tuite[93] compared OS with AS; Patten and colleagues[92] compared OC alone with OC plus OS).
[d] Key biases: partial verification bias; differential verification bias; test review bias; diagnostic review bias.
* Heterogeneity, $p < 0.05$ unless otherwise noted. Where the symbols †, ‡, δ and γ are paired or grouped they indicate that the specificity results are significantly different from each other.

TABLE 12 *MRI: summary results for detection of full-thickness tear*

Outcome subgroup	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
All studies	20	0.89 (0.86 to 0.92)	0.93 (0.91 to 0.95)
By type of MRI[a]			
Conventional	17	0.89 (0.85 to 0.92)*	0.90 (0.87 to 0.93)
Fat-suppressed	5	0.90 (0.83 to 0.94)	0.96 (0.96 to 0.97)*
		NS differences	NS differences
By publication year[b]			
Up to 1994	9	0.94 (0.89 to 0.97)†*	0.93 (0.89 to 0.95)
1995+	11	0.86 (0.81 to 0.90)†	0.94 (0.91 to 0.96)*
		Test for diff,: $p < 0.05$†	NS differences
By age (years)			
<45	4	0.87 (0.74 to 0.94)	0.90 (0.84 to 0.93)†
45–54	11	0.90 (0.86 to 0.94)	0.95 (0.92 to 0.96)*
55+	1	0.67 (0.42 to 0.85)	0.89 (0.56 to 0.98)
Not reported	4	0.92 (0.85 to 0.96)*	0.94 (0.88 to 0.97)†
		NS differences	Test for diff.: $p = 0.04$†
By prevalence (%)			
<40	8	0.90 (0.84 to 0.94)†	0.95 (0.92, 0.96)*
41–60	8	0.93 (0.88 to 0.96)‡*	0.93 (0.89, 0.95)
>60	4	0.76 (0.63 to 0.86)†‡	0.88 (0.77, 0.94)
		Test for diff: $p = 0.01$†; $p < 0.01$‡	NS differences
By reference test			
Arthrography	2	0.92 (0.83 to 0.97)*	0.88 (0.77 to 0.94)†
Arthroscopy	8	0.87 (0.81 to 0.92)	0.94 (0.92 to 0.94)‡*
Surgery (1) or arthroscopy/surgery	7	0.92 (0.86 to 0.95)*	0.92 (0.88 to 0.95)‡*
Other	3	0.85 (0.68 to 0.94)	0.97 (0.90 to 0.99)†‡‡
		NS differences	Test for diff.: $p < 0.01$†‡‡
By views used			
OC, Sag, AX	8	0.91 (0.85 to 0.94)†*	0.94 (0.92 to 0.96)†*
OC, Sg	2	0.94 (0.83 to 0.98)‡δ	0.97 (0.92 to 0.99)†‡
OC, AX	3	0.81 (0.67 to 0.90)†‡	0.92 (0.84 to 0.96)
OC	5	0.78 (0.65 to 0.87)δ	0.91 (0.84 to 0.96)‡
Other (true Sag)	1	100 (0.93 to 1.00)	0.88 (0.76 to 0.94)
Not reported	1	0.91 (0.76 to 0.97)	0.89 (0.76 to 0.96)
		Test for diff.: $p = 0.03$; $p < 0.01$‡; $p = 0.02$δ	Test for diff.: $p = 0.03$†; $p < 0.01$‡
By design			
Prospective	7	0.86 (0.80 to 0.91)†*	0.87 (0.81 to 0.92)†
Retrospective or not reported (3)	13	0.91 (0.87 to 0.94)†	0.95 (0.93 to 0.96)†8
		Test for diff.: $p = 0.03$†	Test for diff.: $p = 0.03$†
Key biases[d]			
1 reported absent	7	0.91 (0.86 to 0.95)*	0.93 (0.89 to 0.95)*
2 reported absent	9	0.88 (0.80 to 0.93)*	0.91 (0.87 to 0.94)*
3 reported absent	4	0.88 (0.81 to 0.93)	0.96 (0.93 to 0.98)
		NS differences	NS differences

[a] Total number of studies adds to 22 because two studies (Sahin-Akyar and colleagues[87] and Reinus and colleagues[86]) compared conventional and fat-suppressed MR

[b] Mean prevalence of tear decreases with each successive group of studies: 37 and 29, respectively.

[c] Sag, sagittal.

[d] Key biases: partial verification bias; differential verification bias; test review bias; diagnostic review bias.

* Heterogeneity, $p < 0.05$ unless otherwise noted. Where the symbols †, ‡ and δ are paired or grouped they indicate that the specificity results are significantly different from each other.

TABLE 13 *MRI: summary results for detection of partial-thickness tears*

Outcome subgroup	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
All studies	14	0.44 (0.36 to 0.51)*	0.90 (0.87 to 0.92)*
By type of MRI[a]			
Conventional	11	0.39 (0.31 to 0.48)*	0.85 (0.81 to 0.89)*
Fat-suppressed	5	0.49 (0.39 to 0.59)*	0.93 (0.89 to 0.95)*
		Test for diff.: $p = 0.05$	NS differences
By publication year[b]			
Up to 1994	5	0.55 (0.43 to 0.66)*	0.90 (0.84 to 0.94)
1995+	9	0.37 (0.29 to 0.46)*	0.90 (0.86 to 0.92)*
		NS differences	NS differences
By age (years)			
<45	4	0.36 (0.25 to 0.48)*	0.84 (0.77 to 0.89)
45–54	7	0.48 (0.38 to 0.58)*	0.93 (0.90 to 0.96)*
55+	0	NA	NA
Not reported	3	Not pooled	Not pooled
		NS differences	NS differences
By prevalence (%)			
<40	6	0.42 (0.33 to 0.52)*	0.92 (0.88 to 0.94)*
40–60	6	0.49 (0.37 to 0.61)*	0.86 (0.81 to 0.90)
>60	2	0.29 (0.13 to 0.53)*	0.91 (0.80 to 0.97)*
		NS differences	NS differences
By views used			
OC, Sag, AX (5) or OC, Sag (2)	7	0.53 (0.43 to 0.62)*	0.91 (0.88 to 0.94)*
OC, AX (3) or OC alone (3)	6	0.31 (0.21 to 0.43)*	0.86 (0.79 to 0.90)
Not reported	1		
		NS differences	NS differences

[a] Total number of studies adds to 22 because two studies (Sahin-Akyar and colleagues[87] and Reinus and colleagues[86]) compared conventional and fat-suppressed MR.
[b] Mean prevalence of tear decreases with each successive group of studies: 33 and 20, respectively.
* Heterogeneity, $p < 0.05$ unless otherwise noted.

The views used in the studies also varied. The oblique coronal (OC) view was utilised in 26 of the 27 studies that provided details, in seven of which it was the only view reported to have been used.[59,63,87–91] Ten studies used the combination of OC, oblique sagittal (OS) and axial (AX) views, two used only OC and OS views, six used only OC and AX views and one used the OS view alone. Two studies compared the accuracy of MRI using two different sets of views. In one,[92] OC alone was compared with OC plus OS, and in the other[93] OC plus OS was compared with OC plus the angled sagittal view.

Outcomes

All of the studies concentrated on the detection of RCTs, with only three examining its performance for the detection of impingement syndrome as a whole.[17,64,88] Twenty studies reported accuracy for the detection of any RCT, 19 for full-thickness tears and 12 for partial tears.

Sample details

The 29 studies included 1633 patients, for a mean sample size of 51. Only five studies included 100 patients or more.[38,59,88,89,93] The mean prevalence of any RCT across the studies was 57% (764 cases and 869 controls). For those that reported it, the mean age of participants was 46 years. Eleven studies had a prevalence of 40% or less and in nine prevalence was 60% or above. All studies except three, two conducted in physical therapy or sports medicine centres[88,94] and one in a general medical centre,[38] were conducted in a hospital radiology or orthopaedics setting, and some clearly had more highly selected populations than others (reflected in prevalence of disease). None of the studies were judged to include an appropriate spectrum of patients. Most did not provide sufficient information on which to judge the spectrum; those including volunteers,[16,17,85] only athletes[94] or where mean age was particularly low,[37,86] were judged not to have included an

appropriate spectrum, although clearly those conducted on athletes will be relevant to those working in sports medicine clinics. A majority of the rest included only those undergoing the reference test and are also unlikely to have included an appropriate spectrum; this included both retrospective studies (13) and those prospective studies where only those already referred for the reference test were included (six).

Reference test

Over half of the studies used an acceptable reference standard: arthroscopy (11 studies), surgery (five studies) or some combination of the two (three studies). In the others the reference test was scored as unacceptable (six studies) or unclear (four studies). One of the latter[88] used the SIT; the others used arthrography as at least part of the reference test. Sufficient details of the technique used were reported in 19 studies. Differential verification, where more than one reference test was used or different areas of the shoulder were investigated using one test, was present in 14 studies. Partial verification bias occurred in 10 studies.

Disease progression bias was a potential problem in six studies where the mean delay between MRI and the reference test was reported to be over 2 months. A further five studies reported that the tests were performed on the same day[71] or within 1 month of each other.[63,95–97] The remaining studies did not report the interval between the tests.

Test interpretation

Twenty-two studies explicitly reported that the MRI diagnosis was reached without knowledge of the reference test result (*Table 10*). As the MRI would almost always be performed before the reference test, it may be safe to assume that those coded as unclear in the quality assessment were in fact blinded to the reference test results. One study[95] reported that the MRI and reference test results were reviewed in conference, and were therefore not interpreted blindly. Diagnostic review bias (knowledge of the MRI result when interpreting the reference test) was very poorly reported. It did not appear to be present in two studies but was reported in four. The remainder did not state whether the MRI results were available to the person interpreting the reference test. Fifteen studies stated that no clinical information was available when conducting the MRI and three stated that it was available.

The criteria used to interpret the MRI scans were all very similar to the original criteria reported by Zlatkin and colleagues[16] and Iannotti and colleagues,[17] although these studies were explicitly referenced by only six of the other studies.[64,86,88,90,92,94,98] Eighteen studies reported that those reading the MRI scans were experienced in the technique: seven reported used musculoskeletal radiologists;[64,82,85,87,94,96,97] five reported length of experience of the readers, ranging from 2 up to 10 years;[37,63,89,92,98] and six gave no further details.[16,17,66,84,88,99]

The studies generally did not discuss the possibility of indeterminate results or how such results had been dealt with. Two reported that radiologists made a 'forced choice' of positive or negative.[92,93] Two others alluded to indeterminate results in their discussion of the criteria used to diagnose tears: one[89] reported that where only very small areas of abnormal signal were identified in the cuff with maintenance of cuff continuity, results were considered negative; the other[90] reported that 'superficial fraying' was included in the 'no tear' category. Withdrawal bias did not appear to be a problem in most of the studies. Withdrawals or losses to follow-up were not reported in the majority. Six studies did report exclusion; in two cases the number and reason for withdrawal were reported and these scored 'no';[71,91] in four cases the number of withdrawals was reported but no reason was given except that test results were lost[87,89] or labral diagnoses were discovered at surgery.[17,85]

Results

The plot of sensitivity against 1 – specificity (false-positive rate) for each study and outcome are presented in *Figure 8*. These indicate a considerable range in both sensitivity and specificity estimates for all outcomes, particularly for the outcomes of any tear and partial-thickness tears. For the detection of full-thickness tears, the points are more clustered together. Spearman correlations between sensitivity and specificity were not significant (*Figure 8*), and there was no obvious trend in the false-positive rates when ordered by the sensitivity rates (*Figure 9*). It was therefore reasonable to attempt a strategy of pooling the sensitivity and specificity rates. Subgroup analyses were conducted to explore reasons for differences between studies but, in general, the heterogeneity remained (see *Tables 11–13*, discussed below).

Sensitivity and specificity

For any RCT, overall pooled sensitivity was 0.83 (95% CI: 0.79 to 0.86) and specificity was 0.86 (95% CI: 0.83 to 0.88), but both were statistically

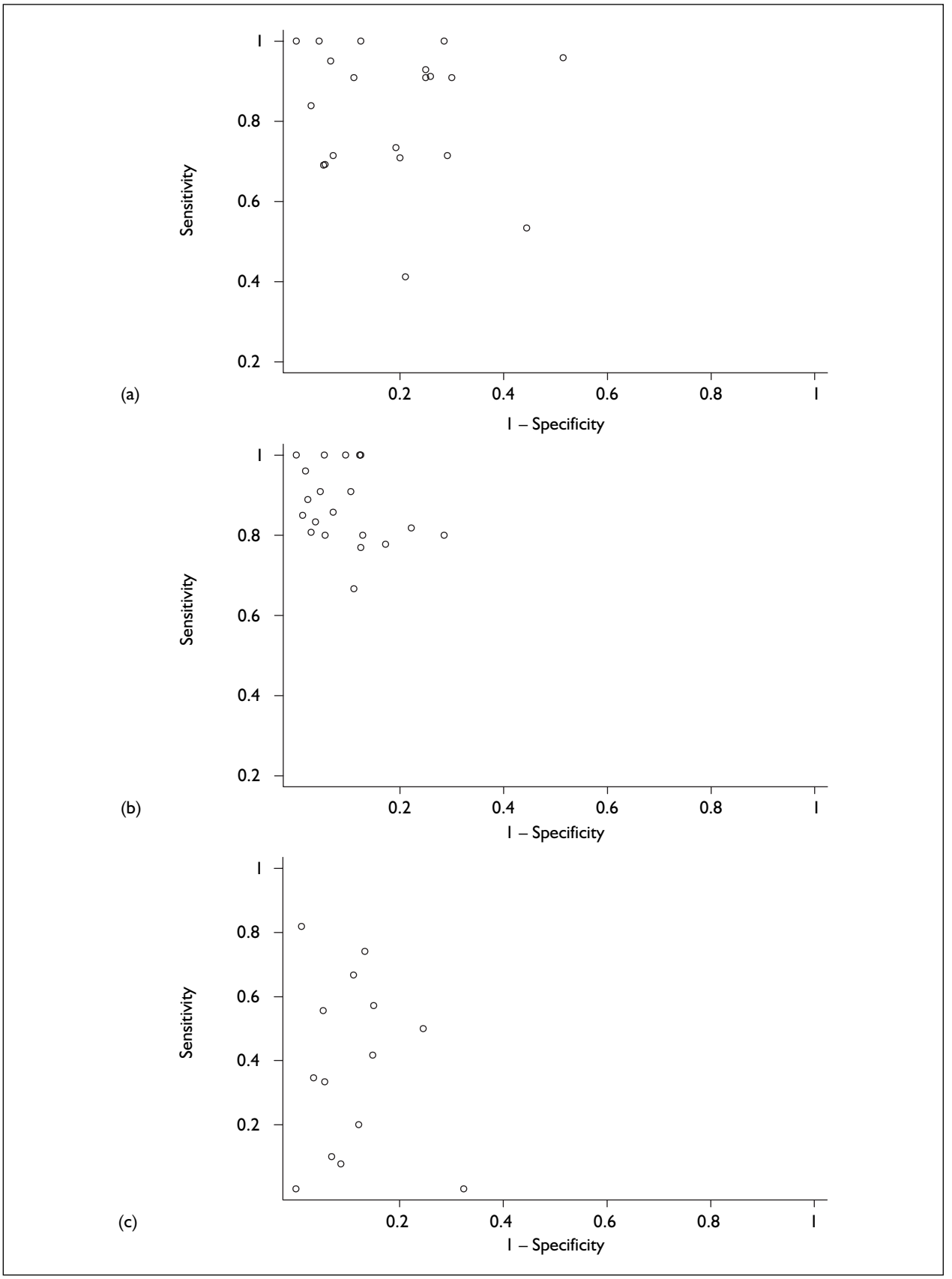

FIGURE 8 *MRI: plot of sensitivity versus 1 – specificity for each outcome. (a) Detection of any tear (*n *= 20): Spearman's rho = –0.07,* p *= 0.76. (b) Detection of full-thickness tears (*n *= 20): Spearman's rho = –0.38,* p *= 0.09. (c) Detection of partial-thickness tears (*n *= 14): Spearman's rho = 0.02,* p *= 0.93.*

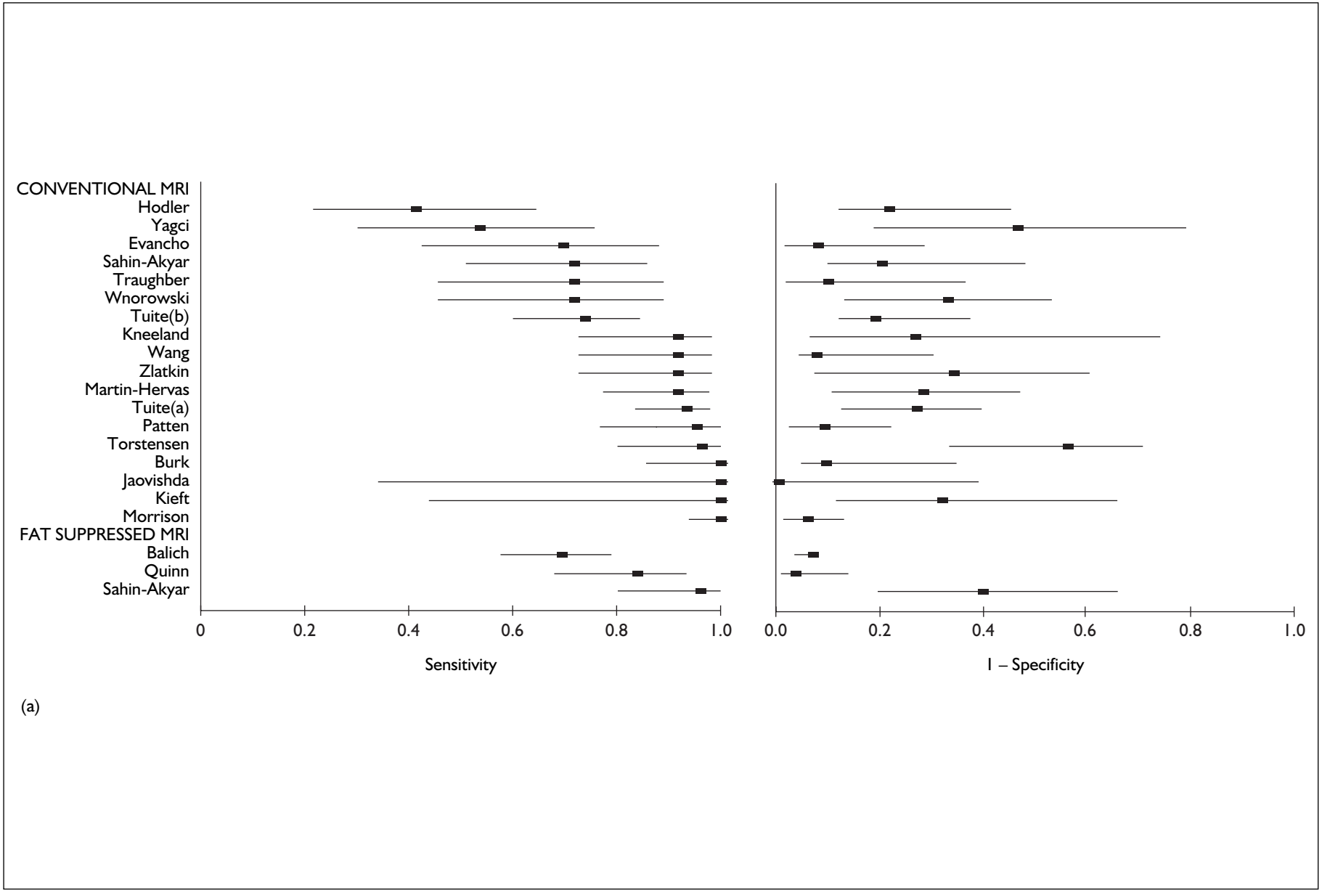

FIGURE 9 *MRI: forest plots of sensitivity against false-positive rate (1 – specificity), per outcome. (a) Detection of any RCT. (cont'd)*

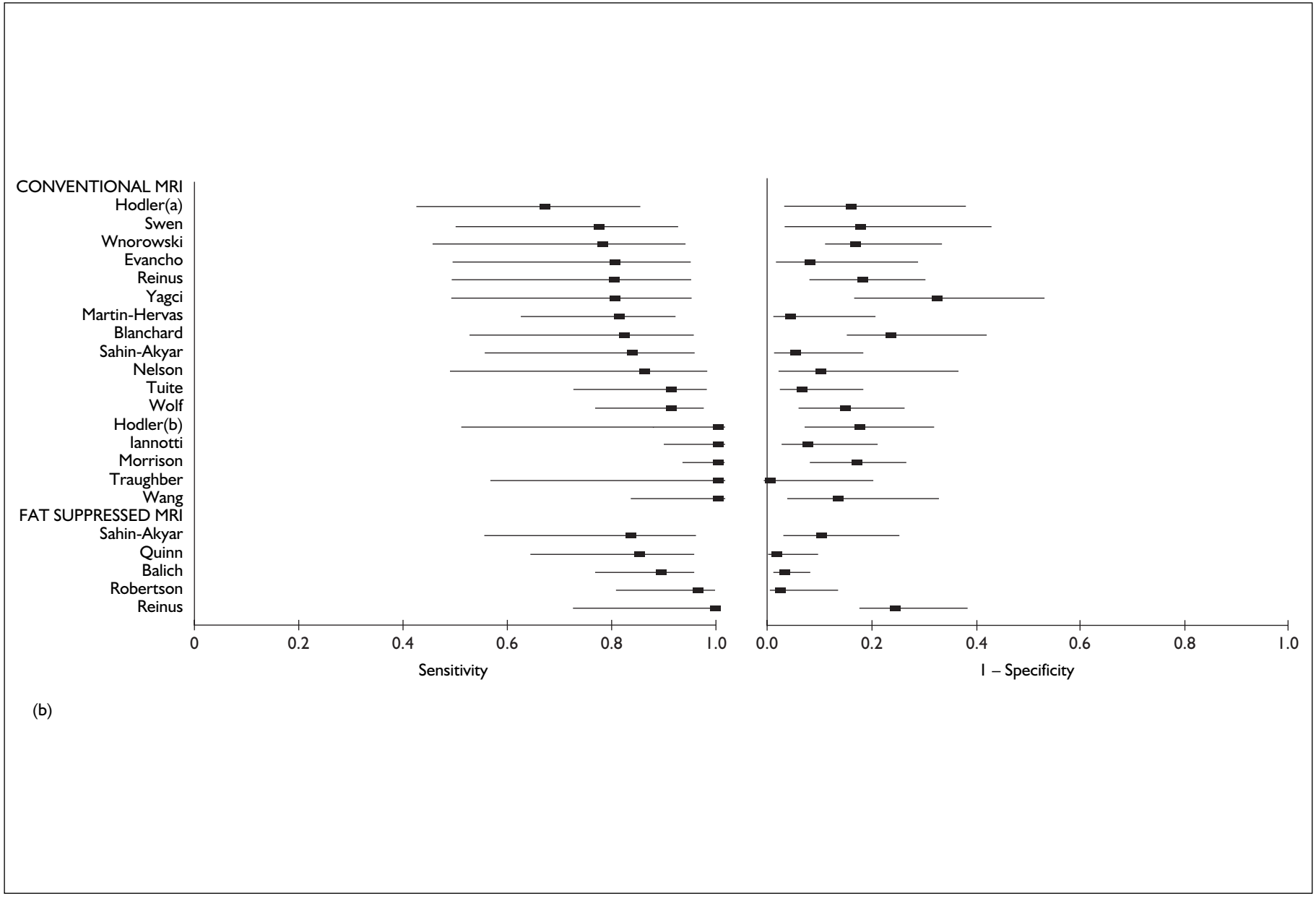

FIGURE 9 *(cont'd)* ***(****b) Detection of full-thickness RCT. (cont'd overleaf)*

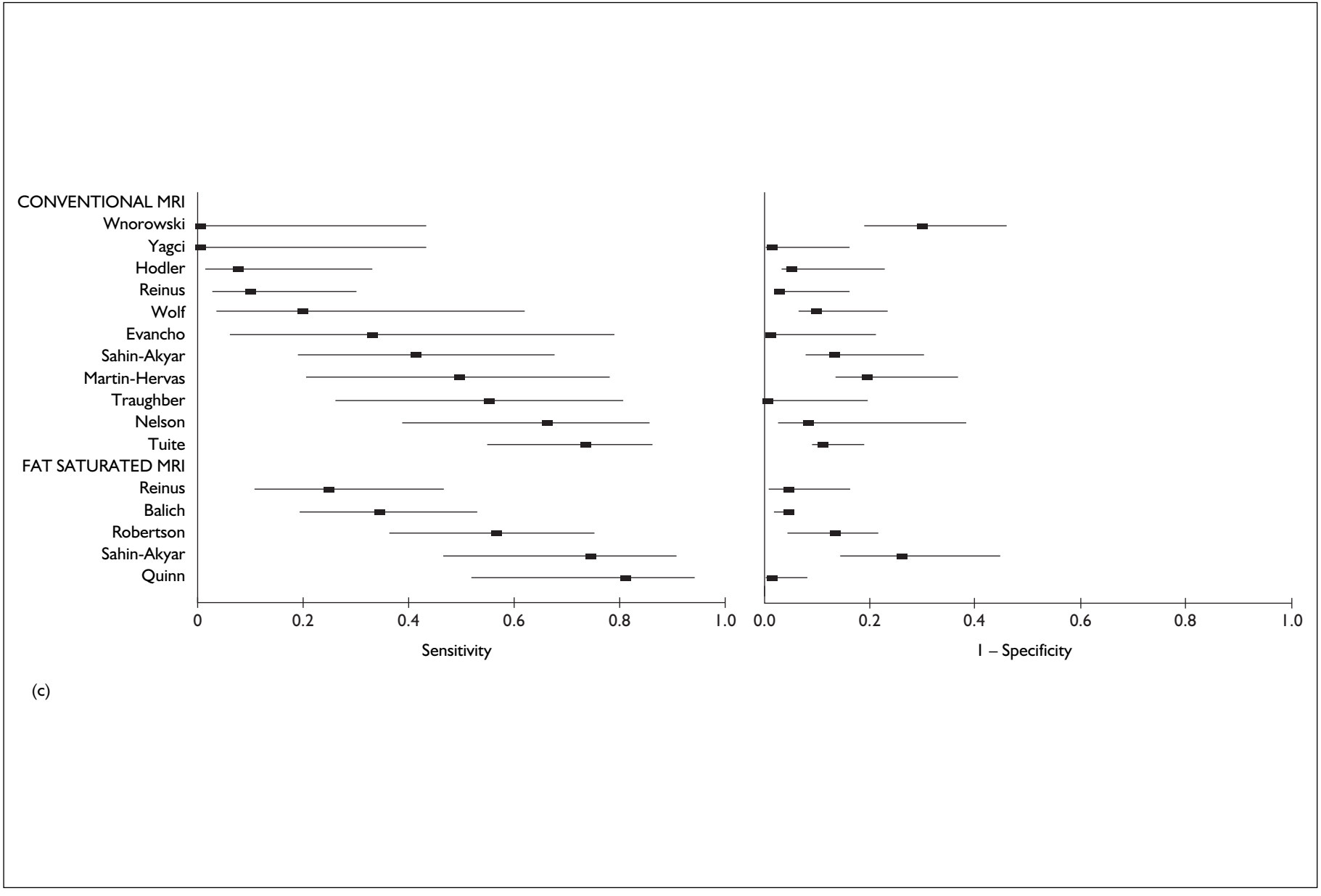

FIGURE 9 *(cont'd) (c) Detection of PT RCT*

heterogeneous with sensitivities ranging from 0.41 to 1.00 and specificities from 0.48 to 1.00. Subgroup analyses to identify possible causes of heterogeneity were not very successful at reducing heterogeneity and few significant differences between subgroups were identified (*Table 11*). The small number of studies in many of the subgroups also needs to be borne in mind when interpreting these results.

Studies using conventional MR produced a higher pooled sensitivity and lower specificity than those using fat-suppressed MR, but the difference was not statistically significant and heterogeneity remained. One of the three studies using the fat-suppressed technique compared conventional with fat-suppressed MR. Sahin-Akyar and colleagues[87] found that the use of the fat-suppressed technique produced higher sensitivity and fairly similar specificity estimates to the conventional technique for the detection of any tear (Appendix 10).

Sensitivity appeared to fall with increasing mean age, although in half of the studies mean age was not reported. Some variation in sensitivity and specificity was found according to the set of views used, but the number of studies per comparison was small. Sensitivity was higher in studies that were prospective in design (0.92, 95% CI: 0.87 to 0.96) compared with those that were retrospective or where the design was not stated (0.80, 95% CI: 0.75 to 0.83), although the difference was not significant. This is in contrast to the result shown for ultrasound (any tear).

For specificities, some significant differences according to prevalence were identified. In those studies with lower prevalence (40% or less), specificity was higher than at higher prevalences: 0.93 (95% CI: 0.89 to 0.95) compared with 0.75 (95% CI: 0.63 to 0.84). Heterogeneity in the latter group was removed but remained in the others.

For full-thickness RCTs, overall pooled sensitivities and specificities were higher than for detection of any tear (*Table 12*), and the studies were not statistically heterogeneous. Again, no significant differences in estimates between studies using conventional MR ($n = 17$) and those using fat-suppressed MR ($n = 5$) were found, although pooled specificity was higher in the latter group.

Sensitivity was unexpectedly higher in low prevalence than high prevalence studies: 0.90 (95% CI: 0.84 to 0.94) compared with 0.76 (95% CI: 0.63 to 0.86), although power was limited, with only four studies in the latter group. The same trend was seen for specificity but the difference between groups was not found to be significant. When studies were subdivided according to year of publication, sensitivity decreased slightly from 94% in studies published before 1994 to 86% in those published after 1995. The concurrent fall in mean prevalence between these two sets of studies (37% versus 29%) probably explains the decrease in sensitivity, and potentially masks any increase in accuracy from improvements in technique or reader experience over time.

Sensitivity and specificity also varied according to the type of views used. The results suggest that accuracy was higher when the OS view was included. This was confirmed in the single study that compared accuracy using the OC view alone with the addition of the OS view; both the sensitivity and specificity were higher using the latter option.[92] Sensitivity also appeared to be lower and specificity higher in prospective studies and in those where a greater number of biases were reported to be absent (*Table 12*).

For detection of partial-thickness RCTs (*Table 13*), the pooled sensitivity estimate was low (0.44, 95% CI: 0.36 to 0.51) although specificity remained high (0.90, 95% CI: 0.87 to 0.92), and studies were again very heterogeneous, especially in terms of sensitivity (see *Figure 9*). Use of fat-suppressed MR did seem to have better accuracy, as did inclusion of the SO view, but sensitivity rates remained low for all subgroups.

TABLE 14 *MRI: summary likelihood ratios*

	No. of studies	Positive LR (95% CI)	Negative LR (95% CI)
Any tear	20	4.85 (3.16 to 7.46)*	0.22 (0.15 to 0.34)*
Full tear	20	10.63 (6.98 to 16.19)*	0.16 (0.11 to 0.23)
Partial tear	14	3.99 (2.34 to 6.80)*	0.66 (0.47 to 0.93)*

* Heterogeneity, $p < 0.01$.

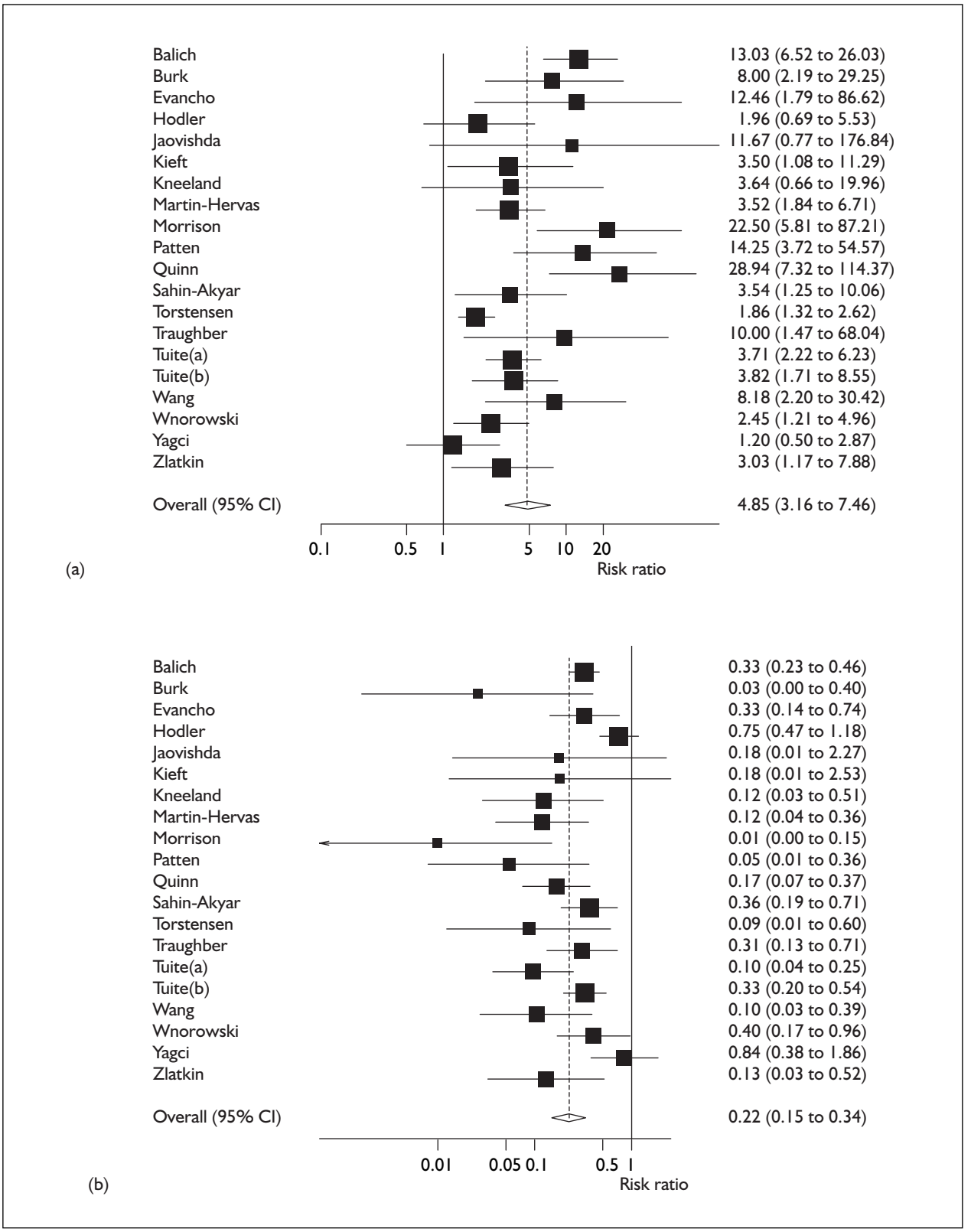

FIGURE 10 *MRI: forest plots of LRs, per outcome. (a) Any tear: positive LR. (b) Any tear: negative LR (cont'd)*

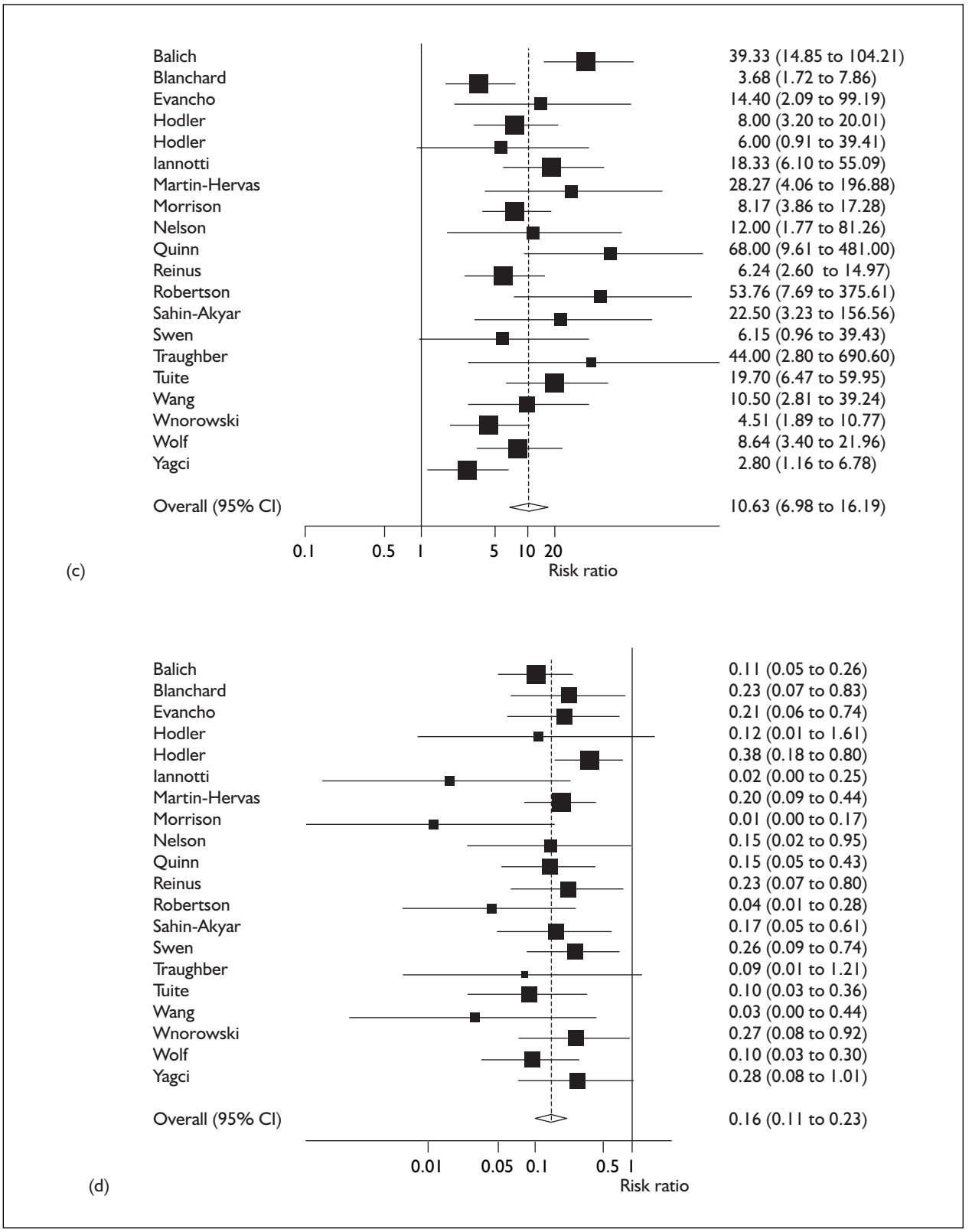

FIGURE 10 *(c) Full tear: positive LR. (d) Full tear: negative LR (cont'd overleaf)*

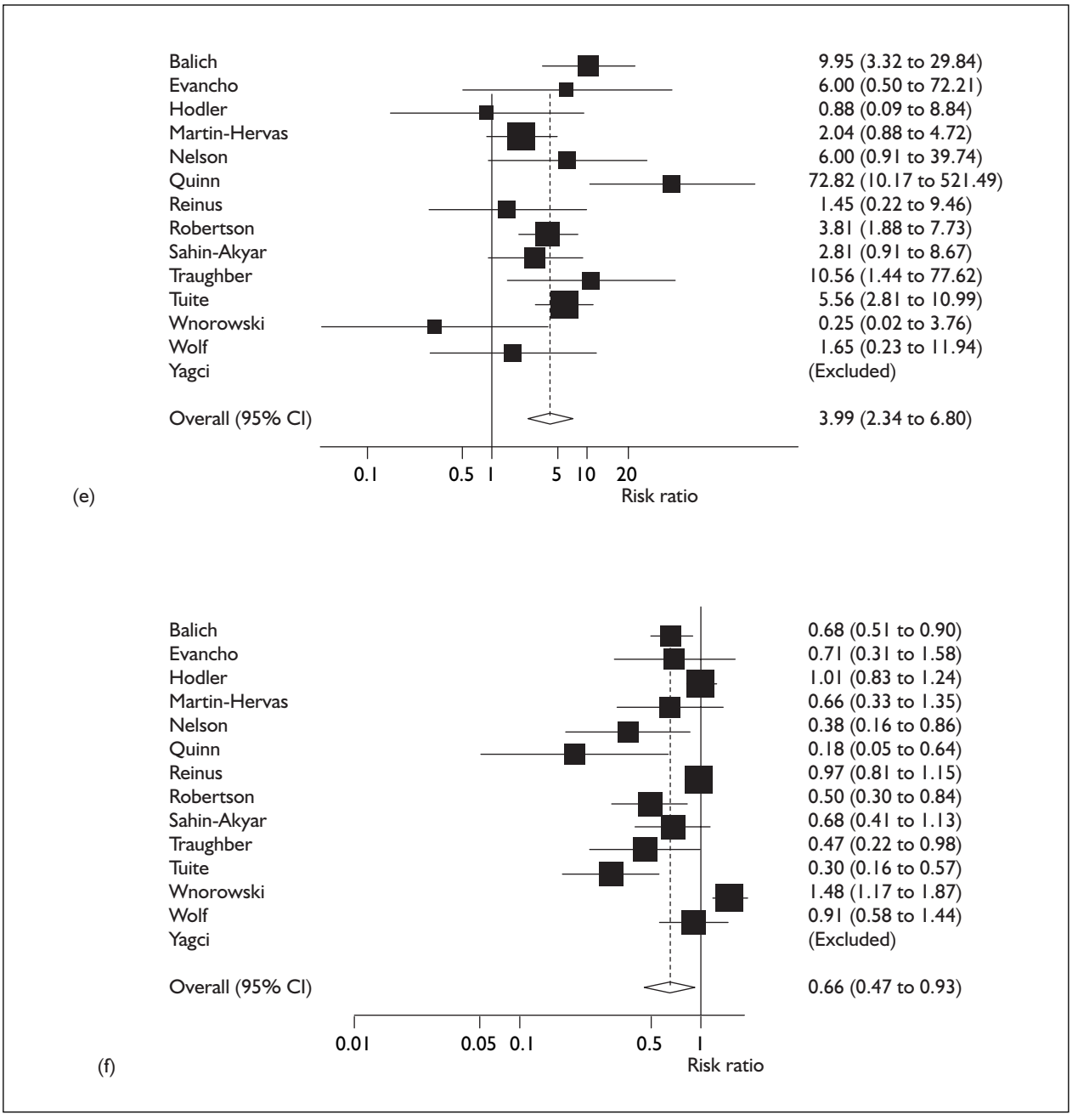

FIGURE 10 *(cont'd) (e) Partial tear: positive LR. (f) Partial tear: negative LR*

LRs

The only homogeneous result in terms of LRs was for detection of a full-thickness RCT: a negative MR result decreases the odds of RCT being present by between 5- and 10-fold (negative LR, 0.16; 95% CI: 0.11 to 0.23; *Table 14*). Negative LRs of between 0.1 and 0.2 can provide strong diagnostic evidence in some contexts, for example, where prevalence is high. *Figure 10* demonstrates that the LRs from all but three of the studies were significant (CIs did not cross one) and all except one[59] included the overall pooled estimate in their CIs.

Studies were heterogeneous for all other pooled LRs, and for partial-thickness RCTs both positive and negative LRs were not impressive (*Table 9*). The forest plots in *Figure 10* show that the positive LRs were low in almost all studies, and only one relatively small study produced a negative LR that would significantly alter the pre-test probability of disease.[84]

Comparisons within individual studies

Eight studies reported accuracy estimates for more than one reader (Appendices 10 and 11). Their

BOX 2 *MRI: impact of a negative test result on the probability of full-thickness tear*

Summary negative LR for full-thickness tear (across 20 studies):		0.16
Pre-test probability of full-thickness tear in these studies (average prevalence):		0.32
Pre-test odds of full-thickness tear	= 0.32/(1 – 0.32)	= 0.47
Post-test odds of full-thickness tear	= 0.47 × 0.16	= 0.075
Post-test probability of full-thickness tear	= 0.075/(1 + 0.075)	= 0.07

A negative MR finding of full-thickness tear decreases the probability of such a tear being present from around 30% to under 10%

results demonstrate that accuracy can vary, even within a single study, and even where all readers are reported to be experienced,[63,86,97] in some cases by a considerable extent. In the two studies where more than two readers with a range of experience were used,[83,85] the most experienced readers had both the highest sensitivity and specificity estimates (Appendix 10). In both studies sensitivity was seen to vary across readers to a slightly greater extent than did specificity, particularly for detection of full-thickness tears. A further study that compared the accuracy of a musculoskeletal radiology fellow with that of a staff radiologist also compared the use of two views: OS versus angled sagittal (AS).[93] The results, in terms of LRs (Appendix 11) indicate that the fellow had similar (low) accuracy with either view, whereas the positive and negative LRs of the staff radiologist both improved with the angled sagittal view.

Discussion

The results overall indicate a similar picture to that seen for ultrasound: a wide variation in accuracy across individual studies for all three outcomes investigated. However, pooled sensitivity and specificity were found to be homogeneous for the detection of full-thickness RCTs, as was the pooled negative LR. Again, similarly to the situation with ultrasound, the criteria for diagnosis of full-thickness tears seem more clearly defined than for partial tears. The actual diagnostic criteria used in the studies were often not reported in detail, but most studies have based their criteria on those developed by Zlatkin and colleagues in the late 1980s.[16] As discussed previously (see the section 'Interpretation of MRI', p. 7), Zlatkin and colleagues did not set out criteria for diagnosis of partial tears.

Focusing on full-thickness tears, when the pooled negative LR is applied to the pre-test probability of disease in these studies (average prevalence across studies) of 32%, the chance of full-thickness tear being present when the MR is concluded to be negative is <7% (*Box 2*). It may be, then, that in some clinical contexts a negative MR finding may be sufficient to rule out the presence of a full-thickness tear. Between-study heterogeneity means that a similar conclusion cannot yet be drawn regarding a positive test result. However, the pooled specificity rate of 0.93 was homogeneous and consistently high across all subgroup analyses, implying that one could be reasonably sure that a positive MR finding is not a false-positive result.

Accuracy of MRI was not greatly affected by the mean age of study participants. However, the results of the subgroup analyses according to prevalence were surprising. Contrary to what was seen for ultrasound, both sensitivity and specificity appeared to be lower in studies with mean prevalence of >60%, although there were only four studies in the latter group. As discussed previously, as full-thickness tears are more common in older people, one might expect studies with higher prevalence of tears to also have higher mean age. One explanation for the trend according to prevalence may be that rather than disease severity increasing with prevalence, the proportion of more 'difficult' cases or 'diagnostic dilemmas' increased. Alternatively, the difference could be spurious or related to variations in quality or lack of power.

The technical aspects of MR also require further investigation. The number of studies using fat-suppressed techniques was low but there is some evidence to suggest that it might increase accuracy, especially for the detection of partial-thickness tears. Routine inclusion of the OS view in the MR sequence may also be beneficial. Although one might expect accuracy to improve over time, with advances in the technique or gained experience in reading the images, we found that sensitivity tended to decrease over time. However, these accuracy estimates are likely to be confounded by decreasing mean prevalence of tears in these studies over time. It is often found that the criteria for selecting patients to undergo a test become less strict as experience with the technique grows and the cost of performing the test fall.

TABLE 15 *MRA: study methods and quality assessment results*

Study	Design[a]	Sample size	Mean age (years)	% male participants	MRI details[b]	Contrast[c]	Ref[d]	Prevalance[e]	Appropriate spectrum[f]	Eligibility criteria stated[f]	Appropriate ref. test[f]	Disease progression bias[f]	Partial verification[f]	Differential verification[f]	Incorporation bias[f]	Details of index test[f]	Details of ref. test[f]	Test review bias[f]	Diagnostic review bias[f]	Clinical info. available[f]	Indeterminate missing[f]	Withdrawal bias[f]
Binkert, 2001[103]	P3	88	51	59	Fat; AX, PS, AC	Gad 2 vs 4 vs Ringer	AS/S	35	?	Y	Y	?	Y	Y	N	Y	N	N	?	N	?	N
Funke, 1996[101]	P2	25	43	64	Con vs Fat OC, OS	Gad	AG	68	?	N	N	N	?	N	N	Y	Y	N	?	N	?	N
Hodler, 1992[90]	?	36	43	67	Con OC, OS	Gad	AS	47	?	N	Y	?	Y ?	N	N	Y	Y	N	?	N	?	N
Loew, 2000[102]	P3	27	48	66	Con 0.2 vs Fat 1.0 T; nr	Gad	AS/S	44[F]	?	?	Y	?	Y	N	N	?	Y	N	?	N	?	N
Pfirrmann, 1999[100]	R	50	51	70	Con; AC, PS, AX vs PS vs AX	Gad	AS/S	42	?	Y	Y	?	N	Y	N	Y	N	N	?	?	?	N
Yagci, 2001[97]	P2	24	52	29	Con; AX, PC, PS	Gad	S/AS	63	?	?	Y	N	N	N	N	Y	N	N	?	N	N	N

[a] Design: P2, prospective, all patients referred for reference test; P3, prospective, selected patients underwent reference test; R, retrospective; ?, design not reported.
[b] MRI details: Con, conventional MRI; Fat, fat-suppressed MRI; OC, oblique coronal view; OS, sagittal oblique view; AX, axial view; AC, angled coronal; PC, paracoronal; PS, parasagittal.
[c] Contrast: Gad, gadolinium contrast agent; Ringer, Ringer solution.
[d] Reference test: AG, arthrography; AS, arthroscopy; S, surgery.
[e] Prevalence (%) of any tear, unless specified: F, full-thickness tear.
[f] Responses to quality assessment criteria: N, no; Y, yes; ?, cannot tell.

TABLE 16 *MRA: summary sensitivity and specificity results*

	No. of studies	Sensitivity (95% CI)	Specificity (95% CI)
Any tear	5	0.88 (0.80 to 0.93)* ($p = 0.06$)	0.83 (0.78 to 0.89)
Full tear	4	0.95 (0.82 to 0.98)	0.93 (0.84 to 0.97)
Partial tear	3	0.62 (0.40 to 0.80)* ($p = 0.05$)	0.92 (0.83 to 0.97)

* Results of heterogeneity test, $p < 0.05$ considered statistically heterogeneous.

Once again, the quality and reporting of the studies are such that the strength of these conclusions is severely limited. For these analyses the number of studies using an inappropriate reference test (arthrography) was low, so the impact of any bias from that source is low. Of particular note is the lack of reporting of indeterminate or unclear results. A certain proportion of such results will occur with any test, especially where there is an obvious subjective component, as with imaging tests. Only four studies described the way in which indeterminate results were dealt with. The way in which exclusion of indeterminate results can inflate sensitivity and specificity in three ultrasound studies is discussed in the results, p. 44.

In conclusion, MRI may be able to rule out full-thickness tears and therefore rapidly identify those in whom surgical treatment may not be beneficial. Its ability to identify correctly full and partial-thickness tears is still up for debate.

MRA

Description and quality of included studies

Six studies investigating the accuracy of MRA for the diagnosis of RC disorders were identified. Summary details of the methods and results are provided in *Tables 15* and *16*. Full study details and results are given in Appendices 12–14.

Interventions

The type of MRI, views and contrast used varied considerably between studies. Five studies[90,97,100–102] used conventional MR, two of which compared results with those using fat-suppressed MR,[101,102] and one used fat-suppressed MR alone.[103] One of the studies comparing conventional and fat-suppressed MR used considerably different MR strengths for each comparison: conventional MR was conducted using a 0.2-T unit, whereas fat-suppressed MR was done using a 1.0-T unit.[102] The views used in the studies also varied. The two earlier studies used the OC and OS views, as was commonly used in the MRI studies.[90,101] Three of the remaining studies used the combination of angled coronal (AC), parasagittal and axial views,[97,100,103] one of which[100] compared results using this set of view to parasagittal alone or axial alone. The remaining study[102] did not report the views used.

All of the studies used a gadoteridol contrast agent for the arthrographic component of the examination, one of which compared a concentration of 2–4 mmol l^{-1} with the use of a Ringer solution.[103]

Outcomes

All of the studies concentrated on the detection of RCTs: five reported accuracy for the detection of any RCT, four for full-thickness tears and three for partial tears (Appendix 13).

Sample details

The six studies included 250 patients, for a mean sample size of 42. The mean prevalence of any RCT across the studies was 45% (113 cases and 137 controls). The mean age of participants was 48. All of the studies were conducted in a hospital radiology and or orthopaedics department. None of the studies provided sufficient information on which to judge the spectrum of patients, but it is unlikely that they will be generalisable outwith these settings.

Reference test

Arthroscopy or surgery were the reference tests used in all studies except one using arthrography alone.[101] Three did not provide sufficient details of the technique used, for example by not stating which areas of the shoulder were examined using the reference test. Differential verification, where more than one reference test was used or different areas of the shoulder were investigated using one test, was present in two studies.[100,103] Partial verification bias occurred in three studies.[90,102,103]

Disease progression bias was definitely not present in two studies[97,101] but was a potential problem in the others where the interval between the tests was not reported.

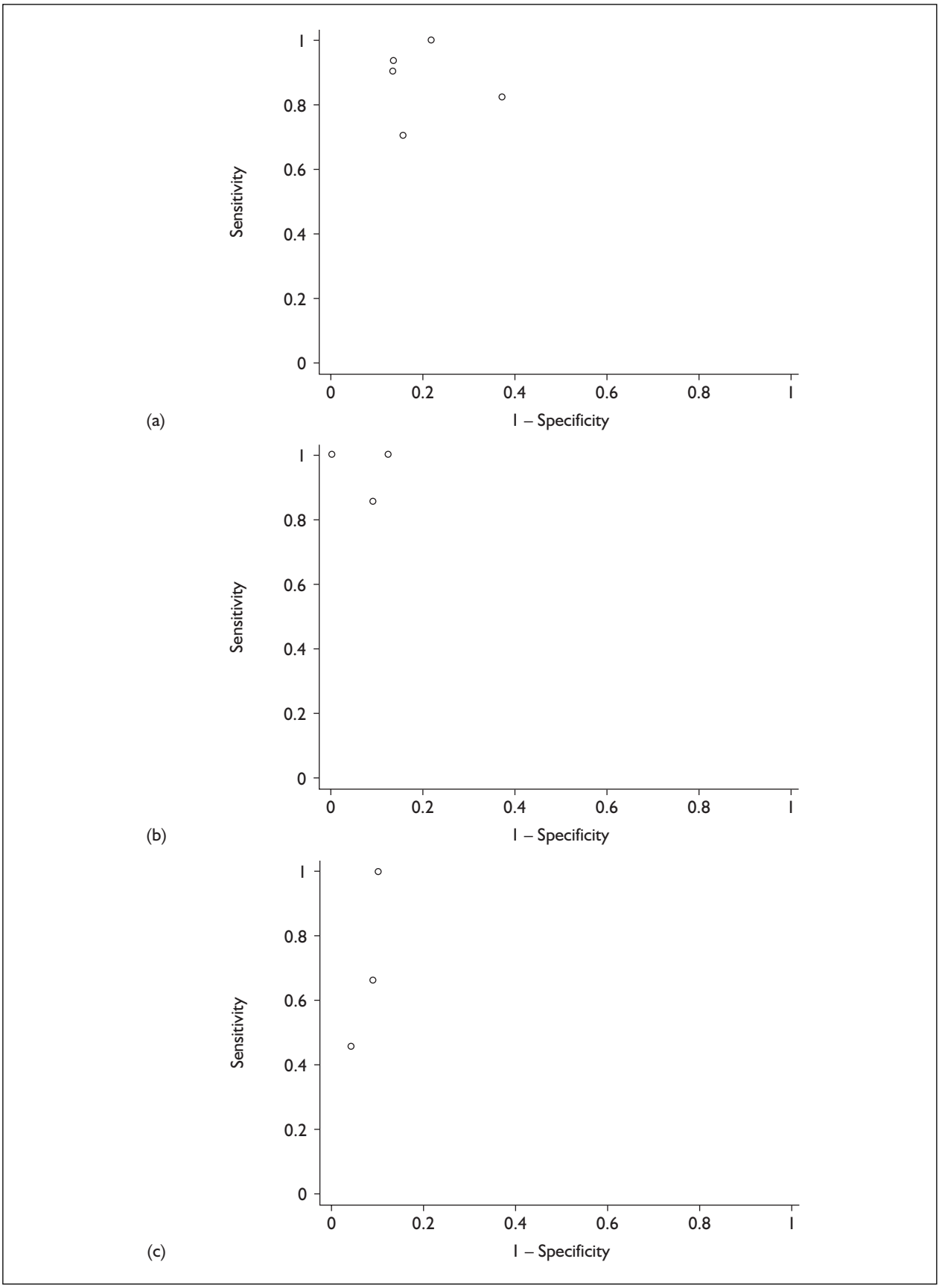

FIGURE 11 *MRA: plots of sensitivity versus 1 – specificity. (a) Detection of any tear* (n = *5): Spearman's rho = 0.10,* p = *0.87. (b) Detection of full-thickness tear* (n = *4): Spearman's rho = 0.27,* p = *0.73. (c) Detection of partial-thickness tear* (n = *3): Spearman's rho = 1.0,* p < *0.01*

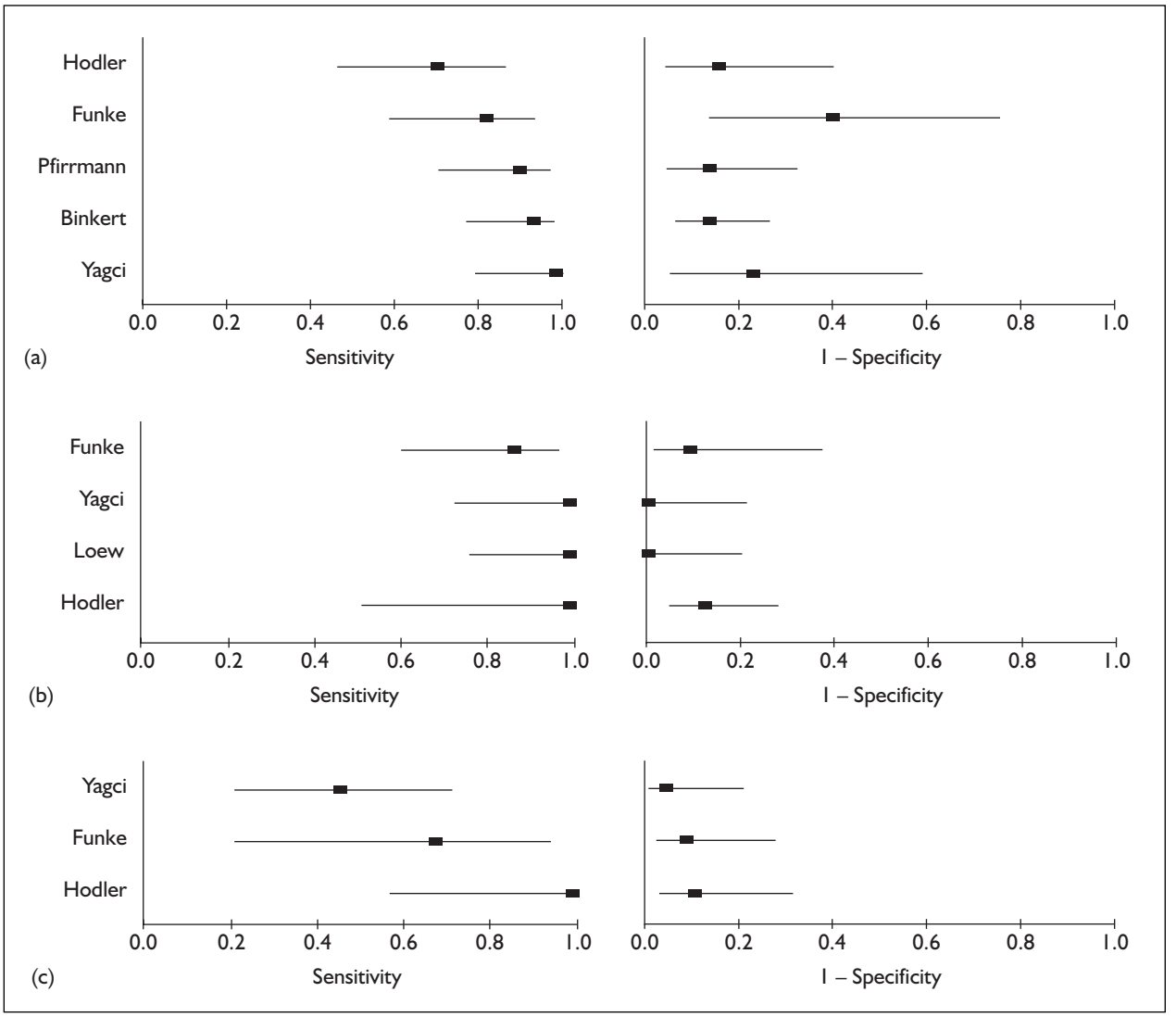

FIGURE 12 *MRA: sensitivity and false-positive rates (1 – specificity). (a) Detection of any RCT. (b) Detection of full-thickness RCT. (c) Detection of partial thickness RCT*

Test interpretation

In all studies the MRA diagnosis was reached without knowledge of the reference test result (*Table 15*). In contrast, none of the studies reported that the reference test was interpreted without knowledge of the MRA result. All of the studies except one[100] explicitly blinded MRA readers to any additional clinical information.

The criteria used to interpret the MRA scans were generally reported, although only one provided any references to earlier studies that had established those criteria.[103] None of the studies discussed the possibility of indeterminate results or how such results had been dealt with. Withdrawal bias did not appear to be a problem in any of the studies.

Results

The plot of sensitivity against 1 – specificity (false-positive rate) for each study and outcome are presented in *Figure 11*. These indicate that for detection of full-thickness tears sensitivity and specificity estimates are quite clustered together but less so for detection of any tear. Spearman correlations between sensitivity and specificity were not significant for these two outcomes, but was highly significant for partial-thickness tears. From the forest plot in *Figure 12(a)* the false-positive rates do appear to increase with the sensitivity rates, though only three studies provided results.

Sensitivity and specificity

The pooled results (*Table 16*) suggest that MRA may be very accurate for the detection of full-

TABLE 17 *MRA: summary likelihood ratio results*

	No. of studies	Positive LR (95% CI)	Negative LR (95% CI)
Any tear	5	4.86 (3.21 to 7.36)*	0.17 (0.07 to 0.38)*
Full tear	4	10.05 (4.71 to 21.46)*	0.11 (0.04 to 0.29)*
Partial tear	3	8.90 (3.64 to 21.78)*	0.43 (0.18 to 1.04)*

* All results statistically homogeneous.

thickness RCTs [overall pooled sensitivity 0.95 (95% CI: 0.82 to 0.98) and specificity 0.93 (95% CI: 0.84 to 0.97), both estimates homogeneous]. Its performance for the detection of partial-thickness tears is less consistent, with some evidence of heterogeneity that may be at least partially due to a threshold effect. The pooled estimates provided should therefore be treated with considerable caution.

No formal subgroup analyses could be performed owing to the small number of studies. All of the results used in the pooled analysis were based on conventional MRI and gadoteridol contrast agent, except those of Binkert and colleagues,[103] who used fat-suppressed MR and three different contrast agents; however, the results did not appear to be very different from those of the other studies (Appendix 13). In the two studies that compared results using conventional versus fat-suppressed MR, one[101] found results to be consistently and considerably better using the fat-suppressed technique, whereas the other[102] found no difference. The former comparison[92] is likely to be confounded by the fact that fat-suppressed images were not interpreted blindly but alongside the standard images. In the latter study,[102] the same results were achieved despite the use of a stronger field strength with the fat-suppressed technique (1.5 T) than the conventional technique (0.2 T).

The study that compared the accuracies of different views[100] generally found better accuracy when all views were combined (AC, parasagittal and transverse views), although one reader did have high accuracy when reading the transverse view alone (Appendix 13). Reading of the three sets of views was conducted by the same radiologists, separated by 3-week intervals, with the full set of views read last; increasing familiarity with the MR reports may have improved accuracy over time.

The study that compared the accuracy using three different contrast agents[103] found little difference in accuracy when either 2 or 4 mmol l^{-1} gadoteridol was used, but sensitivity was higher and specificity slightly lower when the Ringer solution was used, particularly for reader one. As each contrast agent was used in a different group of patients, and furthermore given that the number of patients in each group was small, little weight can be given to these results.

LRs

The pooled LR results are provided in *Table 17* and the corresponding forest plots in *Figure 13*. All pooled results were statistically homogeneous, although any conclusions that can be drawn from the results are at best tentative owing to the low overall number of studies in each comparison.

In terms of full-thickness tears, MRA performs relatively well [overall pooled result: LR positive 10.06 (95% CI: 4.71 to 21.46); LR negative 0.11 (95% CI: 0.04 to 0.29)], although accuracy was lower in the two earlier studies by Funke[101] and Hodler and colleagues.[90] For the detection of any tear, the positive LRs are all <10 for each study, although performance in terms of negative LRs was better for all studies except those of Funke[101] and Hodler and colleagues.[90]

For partial-thickness RCTs, all three studies that presented results produced positive LRs of at least 10, and two[90,97] also produced very low negative LRs.

Discussion

Although the number of studies is small, evidence to date suggests that MRA may be accurate in the detection of full-thickness tears. There is also some suggestion that for detection of partial tears MRA may perform better than ultrasound or MRI, but the available evidence is too sparse to draw more than very tentative conclusions. More evidence from larger prospective studies is required to make a true assessment of its accuracy.

For full-thickness tears, when the pooled positive LR is applied to the pre-test probability of disease

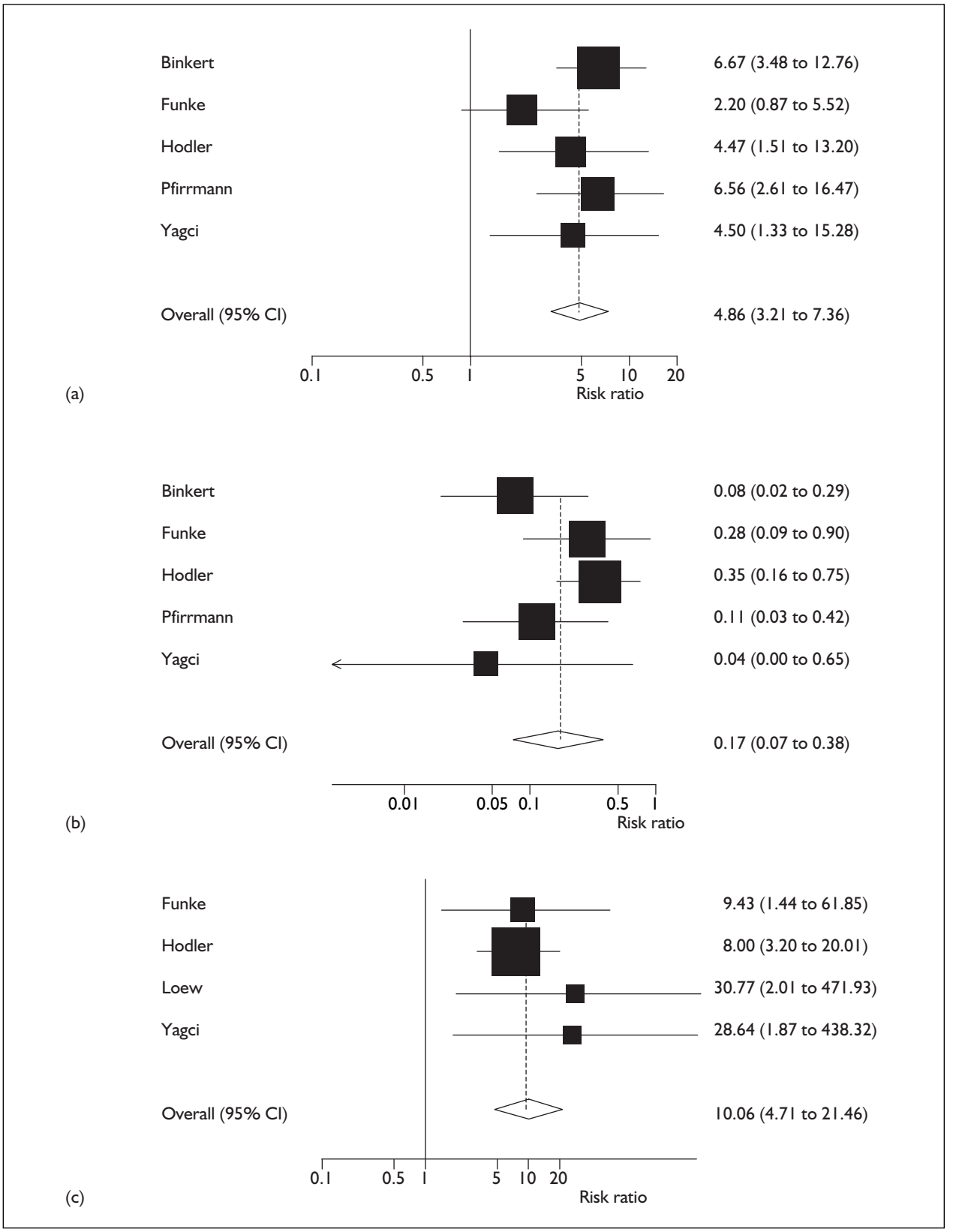

FIGURE 13 *MRA: forest plots of LRs, per outcome. (a) Any tear: positive LR. (b) Any tear: negative LR. (c) FT tear: positive LR (cont'd)*

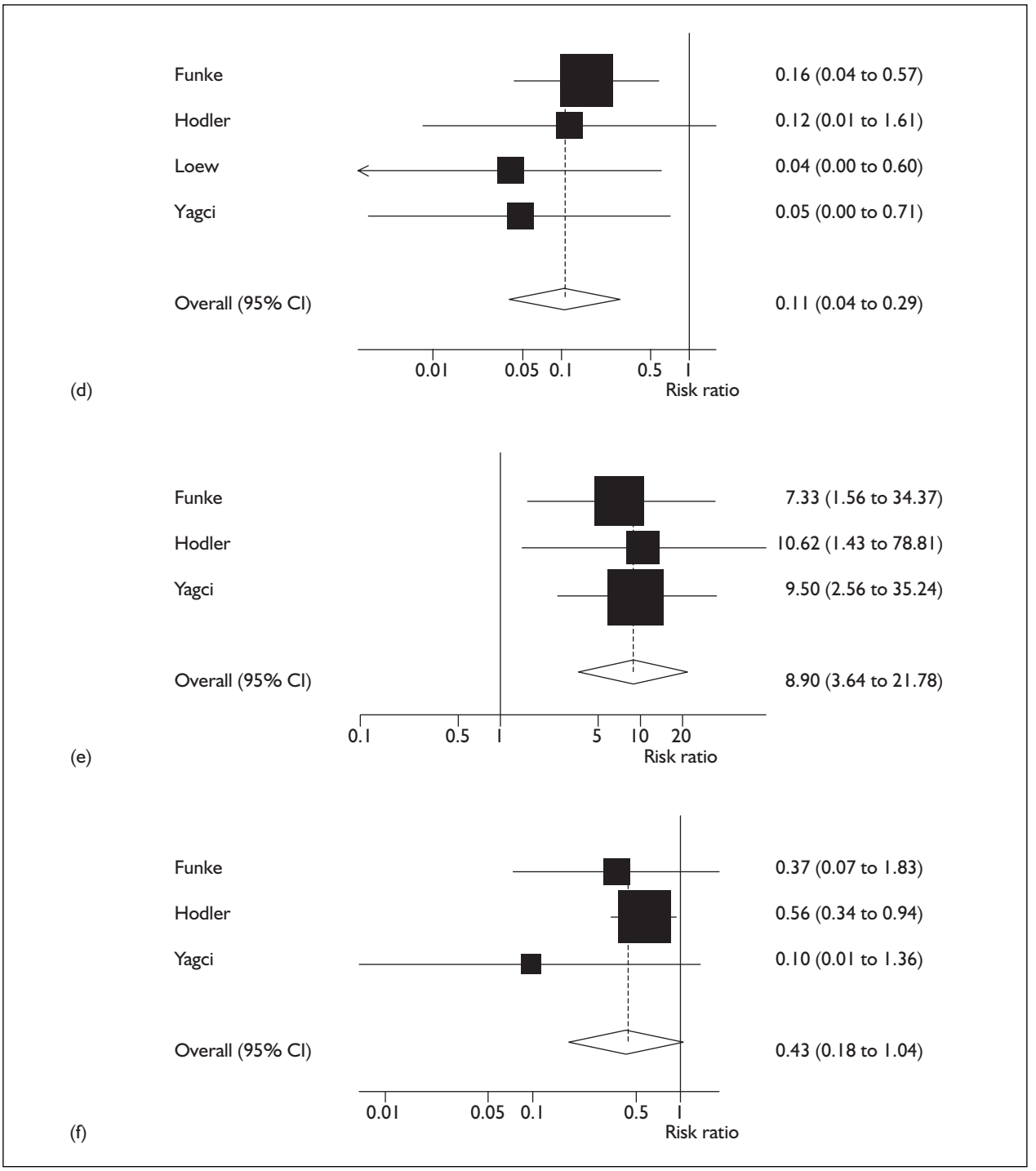

FIGURE 13 *(cont'd) (d) Full-thickness tear: negative LR. (e) Partial-thickness tear: positive LR. (f) Partial-thickness tear: negative LR*

in these studies (average prevalence across studies) of 36%, the chance of a full-thickness tear being present when the MRA is concluded to be positive is over 80%, while a negative MRA reduces the chance of a full tear being present to just over 5% (*Box 3*). There was also some suggestion that the accuracy of the technique has improved over time; however, larger and better quality studies are required to determine the potential contribution that MRA can make to diagnosis.

Disease progression bias, verification bias and possible lack of blinding could all have affected one or more of the studies included. Again, there was a possibility of indeterminate or unclear results in almost all of the studies.

In conclusion, MRA may have some role in the diagnosis of full-thickness and possibly partial-thickness tears, but any such benefit must be set against the invasiveness and potential discomfort

BOX 3 *MRA: impact of a positive or negative test result on the probability of full-thickness tear*

Pre-test probability of full-thickness tear in 4 studies (average prevalence):		0.36
Pre-test odds of full-thickness tear	= 0.36/(1 – 0.36) = 0.5625	
Summary **positive** LR for full-thickness tear:		10.05
Post-test odds of full-thickness tear following positive MRA	= 0.5625 × 10.05 = 5.65	
Post-test probability of full-thickness tear	= 5.65/(1 + 5.65) = 0.85	
A **positive MRA** finding of full-thickness tear increases the probability of such a tear being present from **36%** to over **80%**		
Summary **negative** LR for full-thickness tear:		0.11
Post-test odds of FT tear following negative MRA	= 0.5625 × 0.11 = 0.06	
Post-test probability of full-thickness tear	= 0.06/(1 + 0.06) = 0.06	
A **negative MRA** finding of full-thickness tear decreases the probability of such a tear being present from **36%** to just over **5%**		

to patients from the procedure. Its accuracy needs to be more fully determined before its place in the diagnostic process can be determined.

Studies evaluating more than one test

Ten studies evaluated the accuracy of two different tests against a reference standard. The results of these studies are discussed below and are presented in *Tables 18* and *19*. Full details of the methods, quality and results of these studies can be found in the sections relevant to the individual tests.

Clinical examination versus imaging tests

Three studies evaluated the accuracy of clinical examination and an imaging test.[36–38]

In a study of 42 patients, Read and Perko[36] found clinical examination to be more sensitive but less specific than ultrasound. When translated into LRs, a negative clinical examination could more definitely rule out the possibility of impingement syndrome being present compared with a negative ultrasound result. Positive LRs for both tests were low (*Table 19*).

When compared with MRI, one study[37] found MRI to be much more sensitive than clinical examination, but low specificity estimates meant that both tests performed poorly in terms of LRs. Wolf and Agrawal[38] found a single clinical examination test (the Rent test) to be much more accurate than MRI for the detection of full-thickness RCTs, although only a subgroup of those undergoing clinical examination also underwent MRI (71/109).

Ultrasound versus MRI

Five studies evaluated both ultrasound and MRI: two for the detection of any tear,[66,71] four for the detection of full-thickness tears[59,63,64,66] and two for partial-thickness tears.[64,66]

For the outcome 'any tear', there was some suggestion MRI was more accurate than ultrasound, particularly in terms of negative LRs (a negative MR result would rule out the presence of a tear with more certainty than a negative ultrasound result).

Results for the four studies that presented results for full-thickness tears (in 127 patients) were pooled, and the plot of sensitivity against the false-positive rate is provided in *Figure 14*. Overall, the pooled sensitivity and specificity were almost exactly the same, although this concealed the fact that the MRI estimates were more clustered together than those for ultrasound. In terms of the pooled LRs (*Table 19*), MRI performed slightly better than ultrasound [LR positive 10.4 (95% CI: 4.0 to 26.8) compared with 6.8 (95% CI: 2.8 to 16.1)].

For detection of partial tears, although the accuracy was low, MR was more accurate than ultrasound in both studies.

MRI versus MRA

Two studies evaluated both MRI and MRA. Hodler and colleagues[90] found the accuracy for detection of full-thickness tears to be the same for both tests, but MRA was better at identifying partial-thickness tears (*Tables 18* and *19*). Yagci and colleagues[97] found MRA to perform better than MRI for all outcomes, although the use of fat suppression for MRA may also have contributed to the increased accuracy.

TABLE 18 *Studies comparing accuracy of two tests: sensitivity and specificity results*[a]

Study	Test 1	Se (95% CI)	Sp (95% CI)	Test 2	Se (95% CI)	Sp (95% CI)
	Clinical examination			Ultrasound		
Read, 1998[36]	Approx. 6 signs/symptoms	0.97 (0.85 to 0.99)[I]	0.63 (0.31 to 0.86)[I]	7.5 MHz	0.79 (0.63 to 0.90)[I]	0.88 (0.53 to 0.98)[I]
	Clinical examination			MRI[b]		
Wnorowski, 1997[37]	No details	0.55 (0.28 to 0.79)[A]	0.50 (0.30 to 0.70)[A]	Con; OC, OS, AX; results for CCS readers	0.86 (0.60 to 0.96)[A]	0.52 (0.33 to 0.70)[A]
Wolf, 2001[38]	Rent test	0.96 (0.85 to 0.99)[F]	0.97 (0.89 to 0.99)[F]	nr; nr	0.91 (0.76 to 0.97)[F]	0.89 (0.76 to 0.96)[F]
	Ultrasound			MRI[b]		
Burk, 1989[71]	5 MHz	0.60 (0.36 to 0.80)[A]	0.57 (0.37 to 0.74)[A]	1.5 T; Con; OC, OS, AX	1.00 (0.85 to 1.00)[A]	0.88 (0.64 to 0.97)[A]
Nelson, 1991[64]	5 MHz	0.60 (0.23 to 0.88)[F] 0.36 (0.15 to 0.65)[P]	0.93 (0.69 to 0.99)[F] 0.75 (0.41 to 0.93)[P]	1.5 T; Con; OC, AX	0.86 (0.49 to 0.97)[F] 0.67 (0.39 to 0.86)[P]	0.93 (0.69 to 0.99)[F] 0.89 (0.56 to 0.98)[P]
Hodler, 1991[59]	7.5 MHz	0.93 (0.70 to 0.99)[F]	0.78 (0.45 to 0.94)[F]	1.5 T; Con; OC	0.67 (0.42 to 0.85)[F]	0.89 (0.56 to 0.98)[F]
Martin-Hervas, 2001[66]	7.5 MHz	0.71 (0.54 to 0.83)[A] 0.58 (0.39 to 0.74)[F] 0.13 (0.02 to 0.47)[P]	0.67 (0.48 to 0.81)[A] 1.00 (0.90 to 1.00)[F] 0.68 (0.55 to 0.79)[P]	0.5 T; Con; OC, AX	0.91 (0.77 to 0.97)[A] 0.81 (0.62 to 0.91)[F] 0.50 (0.22 to 0.78)[P]	0.74 (0.55 to 0.87)[A] 0.97 (0.85 to 0.99)[F] 0.75 (0.62 to 0.85)[P]
Swen, 1999[63]	5 and 7.5 MHz	0.92 (0.67 to 0.99)[F]	0.88 (0.53 to 0.98)[F]	1.0 T; Con; OC	0.77 (0.50 to 0.90)[F]	0.88 (0.53 to 0.98)[F]
Pooled estimate (full-thickness tears)		**0.75 (0.62 to 0.84)**	**0.94 (0.85 to 0.98)**		**0.77 (0.65 to 0.86)**	**0.94 (0.85 to 0.98)**
	MRI[b]			MRA[b]		
Hodler, 1992[90]	1.5 T; Con; OC	0.41 (0.22 to 0.64)[A] 1.00 (0.51 to 1.00)[F] 0.08 (0.01 to 0.33)[P]	0.79 (0.57 to 0.91)[A] 0.88 (0.72 to 0.95)[F] 0.91 (0.73 to 0.98)[P]	Con; OC, OS; Gad	0.71 (0.47 to 0.87)[A] 1.00 (0.51 to 1.00)[F] 0.46 (0.23 to 0.71)[P]	0.84 (0.62 to 0.94)[A] 0.88 (0.72 to 0.95)[F] 0.96 (0.79 to 0.99)[P]
Yagci, 2001[97]	1.0 T; Con; PC; PS; AS	0.53 (0.30 to 0.75)[A] 0.80 (0.49 to 0.94)[F] 0.00 (0.00 to 0.43)[P]	0.56 (0.27 to 0.81)[A] 0.71 (0.45 to 0.88)[F] 1.00 (0.83 to 1.00)[P]	Fat; PC; PS; AX; Gad	1.00 (0.80 to 1.00)[A] 1.00 (0.72 to 1.00)[F] 1.00 (0.57 to 1.00)[P]	0.78 (0.45 to 0.94)[A] 1.00 (0.78 to 1.00)[F] 0.89 (0.69 to 0.97)[P]

[a] Se, sensitivity; Sp, specificity. Results presented for: A, any tear; F, full-thickness tear; P, partial-thickness tear; I, impingement syndrome.
[b] MRI and MRA details: Con, conventional MRI; Fat, fat-suppressed MRI; OC, oblique coronal view; OS, oblique sagittal view; AX, axial view; AS, angled sagittal; PC, paracoronal; PS, parasagittal; Gad, gadolinium contrast agent used; CCS, clinical community setting.

TABLE 19 *Studies comparing accuracy of two tests: LRs*[a]

Study	Test 1	LR+ (95% CI)	LR– (95% CI)	Test 2	LR+ (95% CI)	LR– (95% CI)
	Clinical examination			Ultrasound		
Read, 1998[36]	Approx 6 signs/symptoms	2.6 (1.1 to 6.3)[I]	0.0 (0.0 to 0.4)[I]	7.5 MHz	6.4 (1.0 to 40.1)[I]	0.2 (0.1 to 0.5)[I]
	Clinical examination			MRI[b]		
Wnorowski, 1997[37]	No details	1.1 (0.5 to 2.2)[A]	0.9 (0.4 to 2.0)[A]	Con; OC, OS, AX; results for CCS readers	1.8 (1.1 to 2.8)[A]	0.3 (0.1 to 1.0)[A]
Wolf, 2001[38]	Rent test	30.1 (7.7 to 118.0)[F]	0.0 (0.0 to 0.2)[F]	nr; nr	8.6 (3.4 to 22.0)[F]	0.1 (0.0 to 0.3)[F]
	Ultrasound			MRI[b]		
Burk, 1989[71]	5 MHz	1.4 (0.7 to 2.6)[A]	0.7 (0.3 to 1.4)[A]	1.5 T; Con; OC, OS, AX	8.0 (2.2 to 29.2)[A]	0.0 (0.0 to 0.4)[A]
Nelson, 1991[64]	5 MHz	8.4 (1.1 to 63.3)[F] 1.5 (0.3 to 6.1)[P]	0.4 (0.1 to 1.3)[F] 0.8 (0.5 to 1.5)[P]	1.5 T; Con; OC, AX	12.0 (1.8 to 81.3)[F] 6.0 (0.9 to 39.7)[P]	0.2 (0.0 to 0.9)[F] 0.4 (0.2 to 0.9)[P]
Hodler, 1991[59]	7.5 MHz	4.2 (1.2 to 14.4)[F]	0.1 (0.0 to 0.6)[F]	1.5 T; Con; OC	6.0 (0.9 to 39.4)[F]	0.4 (0.2 to 0.8)[F]
Martin-Hervas, 2001[66]	7.5 MHz	2.1 (1.2 to 3.8)[A] 41.3 (2.6 to 660.7)[F] 1.3 (0.9 to 1.8)[P]	0.4 (0.2 to 0.8)[A] 0.4 (0.3 to 0.7)[F] 0.4 (0.1 to 2.5)[P]	0.5 T; Con; OC, AX	3.5 (1.8 to 6.7)[A] 28.3 (4.1 to 196.9)[F] 2.0 (0.9 to 4.7)[P]	0.1 (0.0 to 0.4)[A] 0.2 (0.1 to 0.4)[F] 0.7 (0.3 to 1.3)[P]
Swen, 1999[63]	5 and 7.5 MHz	7.4 (1.2 to 46.5)[F]	0.1 (0.0 to 0.6)[F]	1.0 T; Con; OC	6.2 (1.0 to 39.4)[F]	0.3 (0.1 to 0.7)[F]
Pooled estimate		**6.8 (2.8 to 16.1)[F]**	**0.3 (0.1 to 0.6)[F]**		**10.4 (4.0 to 26.8)[F]**	**0.3 (0.2 to 0.4)[F]**
	MRI[b]			MRA[b]		
Hodler, 1992[90]	1.5 T; Con; OC	2.0 (0.7 to 5.5)[A] 8.0 (3.2 to 20.0)[F] 0.9 (0.1 to 8.8)[P]	0.7 (0.5 to 1.2)[A] 0.1 (0.0 to 0.6)[F] 1.0 (0.8 to 1.2)[P]	Con; OC, OS; Gad	4.5 (1.5 to 13.2)[A] 8.0 (3.2 to 20.0)[F] 10.6 (1.4 to 78.8)[P]	0.3 (0.2 to 0.7)[A] 0.1 (0.0 to 1.6)[F] 0.6 (0.3 to 0.9)[P]
Yagci, 2001[97]	1.0 T; Con; PC; PS; AX	1.2 (0.5 to 2.9)[A] 2.8 (1.2 to 6.8)[F] –[c]	0.8 (0.4 to 1.9)[A] 0.3 (0.1 to 1.0)[F] –[c]	Fat; PC; PS; AX; Gad	4.5 (1.3 to 15.3)[A] 28.6 (1.9 to 438.3)[F] 9.5 (2.6 to 35.2)[P]	0.0 (0.0 to 0.7)[A] 0.0 (0.0 to 0.7)[F] 0.1 (0.0 to 1.4)[P]

[a] LR+, positive likelihood ratio; LR–, negative likelihood ratio. Results presented for: A, any tear; F, full-thickness tear; P, partial-thickness tear; I, impingement syndrome.
[b] MRI and MRA details: Con, conventional MRI; Fat, fat-suppressed MRI; OC, oblique coronal view; SO, sagittal oblique view; AX, axial view; AS, angled sagittal; PC, paracoronal; PS, parasagittal; Gad, gadolinium contrast agent used.
[c] Not estimated for partial tears as sensitivity = 0.

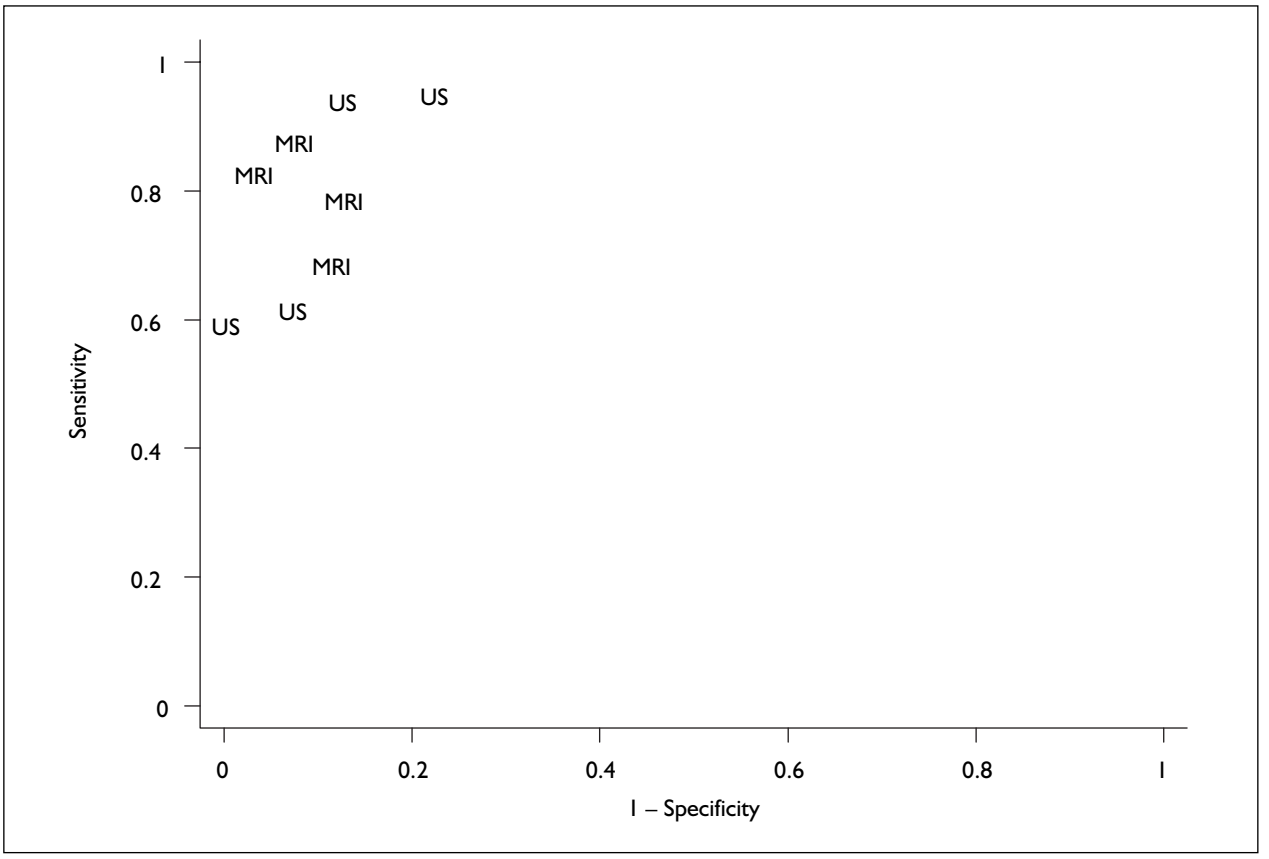

FIGURE 14 *Ultrasound versus MRI: plot of sensitivity against the false-positive rate*

Discussion

Direct evidence for the performance of one test compared with another is very limited. There is some suggestion that MRI may be more accurate than ultrasound and MRA may be more accurate than MRI. However, the individual studies were extremely small and even the pooled analysis for full-thickness tears included only 127 patients. Further research is needed to determine the place of these imaging tests for the diagnosis of RC disorders.

Clinical impact and patient preferences for imaging tests

Diagnostic or therapeutic impact of tests

No studies assessing the impact of ultrasound on diagnostic thinking, therapeutic decisions or patient impact were identified.

One systematic review of studies assessing the effectiveness of MRI of the shoulder was identified.[26] Four primary studies were included in the review;[98,104–106] no additional prospective controlled studies were found. One prospective before and after study examining the impact of MRA on diagnosis and therapeutic decisions was identified.[107] The results of the systematic review and additional MRA study are summarised in *Table 20* and discussed below.

The systematic review reported that MRI can have a wide impact on the clinician's diagnosis and subsequent management plans. Each of the studies effectively compared diagnosis from clinical examination with that following MRI. A new diagnosis was reached in a relatively small proportion of patients in two studies (34 and <15%, respectively) and a much higher proportion in the randomised controlled trial (68%).[106] Confidence in the diagnosis (either increase or decrease) changed in between 24% and 56% of patients. A change in proposed therapy occurred in <15% of patients in one study,[105] but was between 36% and 61% in the remainder. The study comparing outcomes before and after MRA found that a new diagnosis was reached in 57% of cases and a change in management occurred in 49%.[107]

TABLE 20 *Summary of papers examining diagnostic or therapeutic impact of imaging tests*

Study	Design/purpose	Hierarchical level addressed[a]	No. of patients	Diagnoses considered	New diagnosis (%)	Shift confidence (%)	Therapeutic impact[b] (%)	Patient outcomes
Blanchard, 1997[104]	Before/after MRI	3/4/5	71	Various	34	56	61	NS change in SF-36 scores at 6 months
Sher, 1998[105]	Before/after MRI	3/4	65	Various	<15[c]	–	<15	–
Blanchard, 1999[106]	Randomised controlled trial: MRI vs AG	3/4	MRI: 29 AG: 24	Various	MRI: 68 AG: 79	24 33	52 67	–
Blanchard, 1999[98]	Before/after MRI or AG[d]	2/3/4	117	FT RCT		MRI: 30 AG: 37	36 25	–
Zanetti, 1999[107]	Before/after MRA	3/4	73	RC abnormalities	57	incr	49	–

[a] Hierarchical levels at which a test may be evaluated: 2, diagnostic accuracy; 3, diagnostic impact; 4, therapeutic impact; 5, impact on patient outcomes.
[b] Proportion of patients in whom management was changed.
[c] 15% of patients were in 'category one', i.e. a change in either the primary diagnosis or type of treatment; this was not further broken down; 8% of patients were 'category two', i.e. additional clinically relevant findings seen but primary diagnosis and type of treatment were unaltered.
[d] All patients underwent both tests but the order in which they were first presented to the radiologist was randomly determined (MRI first in 46 patients, AG first in 54) and reported results relate to diagnosis following review of that test only.
Adapted from Bearcroft and colleagues.[26]

The authors of the review suggest that the source of patient referrals has a large effect on the size of impact of MRI: when referrals were from a wide variety of sources, or from rheumatologists,[104,106] the therapeutic impact was greater than when referrals came from orthopaedic surgeons. The source of referrals will also impact on the case mix of included patients, which will in turn affect the ability of a test to affect diagnosis and therapy. Differences in sample sizes and study methods will also contribute to the variation in results.

Patient preferences

Only one prospective comparative study of patient preferences was identified; this compared preferences for shoulder MRI or arthrography.[108] The State Trait Anxiety Inventory,[109] a visual analogue scale to assess pain, and a questionnaire to record patients' experiences of the procedures were administered to 88 patients undergoing MRI, 42 undergoing arthrography and 19 undergoing both procedures. Mean anxiety levels were non-significantly higher for those undergoing MRI; a significantly higher proportion of those undergoing arthrography experienced discomfort following the procedure; and a significantly higher proportion of patients found MRI unpleasant or extremely unpleasant compared with those undergoing arthrography (26 versus 7%). Overall, however, the study found no firm patient preference for either investigation.

Cost-effectiveness of alternative diagnostic strategies

Existing cost-effectiveness analyses

Only one cost-effectiveness analysis, conducted in the USA, was identified (full details are provided in Appendix 15). Oh and colleagues[27] presented a decision analytic model to consider the diagnostic choices in a patient with internal derangement of the shoulder and the cost-effectiveness of the diagnostic alternatives. The model compared estimates of the effectiveness and cost-effectiveness of conventional MRI and conventional arthrography (double contrast) with a hypothetical strategy of arthrogram (single contrast) followed by MRA.

The data used in the model were based on clinical experience (prevalence data) and a review of English language publications from 1985 to 1997 (accuracy data). The cost data used were based on 1997 Medicare reimbursement rates. The global imaging costs for conventional MRI, conventional arthrogram and MRA were reported as US$464.57, $163.77 and $628.54, respectively. All patients diagnosed with full or partial RCTs or labral tears were assumed to undergo surgical repair procedures, at costs of $5411, $5324 and $4377, respectively. The measure of effectiveness used was the accuracy of the diagnostic strategy for the diagnosis of each type of tear.

The base case analysis reported the average effectiveness of conventional arthrogram, conventional MRI and MRA (as adjunct therapy) to be 0.6610, 0.6715 and 0.7204, respectively. We have interpreted these cost-effectiveness results as cost per accurate diagnosis, although this was not explicitly stated by the authors. The average costs for these three diagnostic strategies were estimated to be $1090, $2033 and $2339, respectively. The costs for diagnosis and surgical intervention have been weighted in the cost-effectiveness decision model, but the cost breakdowns/components were not presented. The marginal cost-effectiveness was found to be:

- $89,895 for conventional MRI compared with conventional arthrogram
- $21,029 for MRA as adjunct therapy compared with conventional arthrogram
- $6250 for MRA compared with MRI.

Cost of imaging tests

Data from the Finance Department at Southampton General Hospital put the direct costs of each imaging procedure as follows (2000–1 prices):

- ultrasound: £44
- MRI: £101
- MRA: £141.

The last two costs will be more sensitive to the volume of patient throughput than will ultrasound, owing to the higher capital cost of MR imagers.

Chapter 5
Discussion and conclusions

Summary of key findings

Only a small number of studies of clinical examination were identified – too few studies to draw any conclusions regarding individual clinical examination tests. The meta-analysis suggests that clinical examination as a whole, when carried out by relatively specialised clinicians such as orthopaedists, may be useful at ruling out RCTs (high sensitivity and negative LR) but less accurate at detecting such tears when they are present (low specificity and positive LR). 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. No study was reported to have been carried out in a primary care setting; however, accuracy could be expected to be lower given the wider spectrum of conditions encountered, lower prevalence of RCT, and lower level of specialism in shoulder examination.

A large number of studies evaluating ultrasound and MRI were identified, and a small number that had examined MRA. As with the clinical examination studies, only a very small minority of the imaging studies examined the ability of these tests to detect RC **disorders** as opposed to RC **tears**. Although it has been estimated that up to 90% of shoulder pain is due to soft tissue disorders, we do not really have any indication of what proportion of this is down to tears as opposed to other soft tissue lesions, or how good the available tests are at identifying these other lesions.

For most of the three outcomes that could be investigated (detection of any RCT, full tear or partial-thickness tear), sensitivity and specificity estimates across studies for all imaging tests were highly heterogeneous. For each of the tests, results were most homogeneous for the detection of full-thickness RCTs. In each case this seems to be because the criteria for diagnosis of full-thickness tears are more clearly defined than for partial tears. Despite the heterogeneity, the pooled results for each test were remarkably similar to each other. For example, for full-thickness tears, pooled sensitivity for ultrasound, MRI and MRA were 87%, 89% and 95% and specificity 96, 93 and 93%, respectively. The four studies directly comparing MRI and ultrasound also found accuracy to be similar for the two tests.

In terms of LRs, all of the imaging tests may provide convincing evidence of the presence or absence of a full-thickness tear. When the LRs are applied to the pre-test probability of disease, the likelihood of such a tear being present (or absent) is increased (or decreased) markedly. Whether a single positive or negative test result is sufficient to select those who may be potential candidates for surgical treatment without further investigation remains to be determined. It would appear that on the whole a positive test result may be a sufficiently reliable indicator of full-thickness tear, at least in a spectrum of patients similar to those enrolled in the studies, but a negative result (on any of the tests) is not sufficient to rule out the presence of a tear, especially a partial-thickness tear.

All of the imaging tests were less accurate for partial-thickness tears but indirect comparison between tests suggests that MRA and ultrasound might be more accurate at detecting such tears than MRI (positive LRs of ~9). It is worth noting that the clinical significance of some partial tears (for example, those identified by 'minimal fraying' of the RC) remains to be determined. Some tears such as intrasubstance RCTs would be unlikely to be repaired even when detected, although, indeed, not all patients with full-thickness tears will undergo surgery, particularly if they do not experience sufficiently severe symptoms.

Explanations for variation among study results

Very little explanation was found for the heterogeneity seen between studies, but this could largely be as a result of the poor quality and poor quality of reporting of the studies. Despite the relatively large number of studies identified for ultrasound and MRI, the number of studies in some of the subgroups examined was small, and the power of the analyses was therefore limited.

Furthermore, apparent trends in accuracy were not consistent across tests. For example, in both groups of studies, accuracy appeared to increase with increasing mean age. As full-thickness tears tend to be more common in older people, one would expect accuracy also to increase with prevalence. However, sensitivity was found to increase with increasing prevalence in the ultrasound studies and to decrease with increasing prevalence in the MRI studies. The former result is consistent with what one would expect to see if spectrum bias were present: studies conducted in samples with a high prevalence of RCT are likely to have included a higher proportion of more severe cases that are more easily discernible on imaging tests. The trend with prevalence in the MRI studies is more difficult to explain. The case mix, or spectrum, of included patients was almost never reported in any detail, but the eligibility criteria used in many of the studies imply that the samples were not representative of the more general population who experience shoulder pain and who might undergo these tests in practice. If this is true, the accuracy of these tests will not be the same as demonstrated here.

The inappropriate use of arthrography as a reference test probably had a limited impact on the results of most of the analyses owing to the small number of studies using arthrography that were included in each analysis. The impact of using an imperfect reference test varies according to the prevalence of disease in a given sample:[81] where prevalence is low, sensitivity will tend to be underestimated and where prevalence is high, it is specificity that is underestimated. Where studies with varying prevalences are pooled, the effect on the overall pooled estimate will vary. In this review, the use of arthrography could have had the most impact on the analysis of the ability of ultrasound to detect any RCT, where 10 out of 29 studies used arthrography alone. In these studies, sensitivity seemed to be underestimated in comparison with those using other reference tests, but the differences were not statistically significant.

Partial and/or differential verification bias was a potential problem in a large number of studies. Most studies included only those patients who underwent the reference test and the total number of patients in the population who underwent ultrasound, MRI or MRA was left unreported. Each of these factors has the potential to inflate accuracy, as does the lack of blinding when interpreting the index or reference test result. If the result of one test is known when interpreting the other, the reader might spend more (or less) time searching for a tear if they know that another test has (or has not) already found one. This would tend to increase the agreement between the results of the two tests and also increase accuracy. The subgroup analyses for ultrasound showed that sensitivity was lower where two or more of these biases were reported to be present or not reported compared with those where they were absent, although both subgroups of studies remained heterogeneous. The MRI studies showed little variation according to whether key biases were reported to be absent.

Of further note is the lack of reporting of indeterminate or unclear results. A certain proportion of such results will occur with any test, especially where there is an obvious subjective component, as with imaging tests. As discussed in previous chapters, only five ultrasound studies and no MRI studies mentioned the number of indeterminate results found. Exclusion of these results can markedly increase sensitivity and/or specificity.

Between-study differences in the tests themselves may also have affected test accuracy. For example, it might be expected that higher frequency ultrasound scanners increase accuracy by providing the sonographer with a better quality image. There was some suggestion that this occurred, but none of the studies compared results using different frequencies and several authors increased the frequency of scanner used during the course of their study. For MRI, it has been suggested that the use of fat suppression increases accuracy. Again, our results seemed to support this hypothesis, but the differences were not significant and too few studies have compared the conventional and fat-suppressed techniques to allow a true comparison to be made.

Our results also highlight the variation in accuracy between different readers, even where all readers were reported to be 'experienced'. Those who compared the results of readers with different levels of experience did find accuracy to be highest among the most experienced readers, as would be expected. It is possible that results achieved by radiologists engaged in studies of imaging tests may not be transferable to usual practice, partly because readers may be more attentive when engaged in a clinical study and partly because those involved in studies may not be representative of the wider community of radiologists.

Very few studies have gone beyond trying to establish the effect of these tests on outcomes

other than accuracy. Design and quality differences mean that as yet no conclusions can be drawn regarding the ultimate impact of imaging tests for shoulder pain.

Strength and limitations of our review

The systematic nature of the review means that we are likely to have identified the majority of the published studies. The literature search was comprehensive, using a wide range of electronic databases and relatively broad search terms, such that all of the indexed literature should have been picked up. Two reviewers were involved at every stage in the review procedure, such that mistakes due to human error should be limited. A quality assessment tool that has been developed according to scale development principles was adapted and applied to each of the included studies.

Empirical evidence suggests that studies with significant or favourable results are more likely to be published than those with non-significant or unfavourable results.[110] There is as yet no evidence for the degree of publication bias likely in the field of diagnostic tests, but there is no reason to believe that it will be any better than in studies of therapeutic interventions. It is therefore possible that we have missed a proportion of English language studies. Time and resource constraints meant that we could not assess any foreign language papers for inclusion.

Nevertheless, we did identify a large number of studies, particularly for ultrasound and MRI. The majority of studies were small and their results heterogeneous – it is unlikely that the inclusion of additional studies, if similar to those identified, would have significantly changed our results.

Recommendations for research

1. A large prospective follow-up study of patients with shoulder pain in primary care is needed to inform our understanding of the natural history and epidemiology of shoulder pain – diagnoses, therapeutic decisions and referrals should be recorded and long-term follow-up of all patients implemented. Remaining questions include:
   (a) What proportion of shoulder pain (in the population and in primary care, rather than secondary care) is caused by RCTs (partial and full thickness)?
   (b) What is the natural history of partial tears and full-thickness RCTs?
   (c) What are the outcomes in terms of pain, function, surgery, physiotherapy and cost in patients with these conditions who have/have not had imaging?

2. For those patients referred to secondary care, a prospective cohort study of clinical examination, ultrasound and MRI, alone and/or in combination, is also needed. The ability of these tests not only to diagnose soft tissue shoulder disorders but also to inform treatment decisions needs to be determined:
   (a) Diagnostic accuracy must be established in a wide spectrum of patients, against an appropriate reference test, and avoiding the major sources of bias such as verification bias, lack of blinding and inclusion of all indeterminate results. An appropriate reference test in this instance is arthroscopy alone.
   (b) The extent to which diagnostic tests ultimately inform and affect patient management and outcomes should be considered. This can be built into a diagnostic accuracy study by recording any changes in diagnosis and/or treatment decisions after each subsequent test is interpreted. Follow-up of patients to determine long-term outcome would be possible but difficult to attribute to the use of any one test or even diagnostic strategy. If there is doubt over the most appropriate treatment, given a diagnosis, a treatment arm could be considered for inclusion in the study.

   The effect of ultrasound and/or MRI can be examined by all patients undergoing both tests and randomising the order in which they are received.

3. Various aspects of the tests may also warrant further investigation, for example:
   (a) performance of clinical examination by less specialist clinicians or physiotherapists
   (b) use of fat-suppressed versus conventional MRI
   (c) use of 12–15 MHz ultrasound transducers as opposed to 10-MHz or lower instruments.

Conclusions

Our results suggest that clinical examination by specialists can rule out the presence of an

RCT, and that either MRI or ultrasound could equally be used for detection of full-thickness RCTs. Although still not by any means accurate, ultrasound may be better at picking up partial tears. Given the large differential in the cost of the two procedures, the implication from current evidence is that ultrasound is the more cost-effective test to use in a specialist hospital setting for identification of full-thickness tears. Whether or not these results are transferable to settings with lower prevalence, different spectrum of disease and less specialised clinicians, such as in primary care, remains to be determined.

Acknowledgements

The authors wish to thank their expert advisers for their advice and comments on the draft protocol and draft report: Dr Phillip Bearcroft, Consultant Radiologist, Department of Radiology, Addenbrooke's Hospital, Cambridge; Professor William J Gillespie, Dean, Hull York Medical School; and Dr Karen Walker-Bone, Senior Research Fellow and Honorary Consultant Rheumatologist, MRC Environmental and Epidemiology Unit, Southampton General Hospital.

They would also like to thank Liz Payne for conducting the literature search and Jonathan Shepherd and Colin Green for assistance with data extraction. Thanks are also due to Julie Renfrew, Management Accountant at Southampton General Hospital, for providing cost data.

This report was commissioned by the NHS R&D HTA Programme. This report remains the responsibility of the Southampton Health Technology Assessments Centre, Wessex Institute for Health Research and Development, University of Southampton, and the views expressed are those of the authors and not necessarily those of the NHS R&D Programme. Any errors are the responsibility of the authors.

References

1. Urwin M, Symmons D, Allison T, Brammah T, Busby H, Roxby M, *et al*. Estimating the burden of musculoskeletal disorders in the community: the comparative prevalence of symptoms at different anatomical sites, and the relation to social deprivation. *Ann Rheum Dis* 1998;**57**:649–55.
2. Chard MD, Hazleman R, Hazleman BL, King RH, Reiss BB. Shoulder disorders in the elderly: a community survey. *Arthritis Rheum* 1991;**34**:766–9.
3. Bongers PM. The cost of shoulder pain at work. *BMJ* 2001;**322**:64–5.
4. Speed C, Hazelman BL. Shoulder pain. *Clin Evid* 2001;**5**:850–64.
5. Croft P, Pope D, Silman A. The clinical course of shoulder pain: prospective cohort study in primary care. Primary Care Rheumatology Society Shoulder Study Group. *BMJ* 1996;**313**:601–2.
6. de Winter AF, Jans MP, Scholten RJ, Deville W, van Schaardenburg D, Bouter LM. Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. *Ann Rheum Dis* 1999;**58**:272–7.
7. McMahon PJ, Sallis RE. The painful shoulder. Zeroing in on the most common causes. *Postgrad Med* 1999;**106**:36–8, 41–3, 47–9.
8. Boenisch U, Lembcke O, Naumann T. Classification, clinical findings and operative treatment of degenerative and posttraumatic shoulder disease: what do we really need to know from an imaging report to establish a treatment strategy? *Eur J Radiol* 2000;**35**:103–18.
9. Zanetti M, Hodler J. Imaging of degenerative and posttraumatic disease in the shoulder joint with ultrasound. *Eur J Radiol* 2000;**35**:119–25.
10. Green S, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. *BMJ* 1998;**316**:354–60.
11. van der Windt DA, Koes BW, Boeke AJ, Deville W, De Jong BA, Bouter LM. Shoulder disorders in general practice: prognostic indicators of outcome. *Br J Gen Pract* 1996;**46**:519–23.
12. Winters JC, Sobel JS, Groenier KH, Arendzen JH, Meyboom de Jong B. The long-term course of shoulder complaints: a prospective study in general practice. *Rheumatology* 1999;**38**:160–3.
13. Guckel C, Nidecker A. Diagnosis of tears in rotator-cuff-injuries. *Eur J Radiol* 1997;**25**:168–76.
14. Woodward TW, Best TM. The painful shoulder; part I. Clinical evaluation. *Am Fam Physician* 2000;**61**:3079–88.
15. Lee SU, Lang P. MR and MR arthrography to identify degenerative and posttraumatic diseases in the shoulder joint. *Eur J Radiol* 2000;**35**:126–35.
16. Zlatkin MB, Iannotti JP, Roberts MC, Esterhai JL, Dalinka MK, Kressel HY, *et al*. Rotator cuff tears: diagnostic performance of MR imaging. *Radiology* 1989;**172**:223–9.
17. Iannotti JP, Zlatkin MB, Esterhai JL, Kressel HY, Dalinka MK, Spindler KP. Magnetic resonance imaging of the shoulder. Sensitivity, specificity, and predictive value. *J Bone Joint Surg Am* 1991; **73**:17–29.
18. van der Heijden GJ, van der Windt DA, de Winter AF. Physiotherapy for patients with soft tissue shoulder disorders: a systematic review of randomised clinical trials. *BMJ* 1997;**315**:25–30.
19. van der Heijden GJ, van der Windt DA, Kleijnen J, Koes BW, Bouter LM. Steroid injections for shoulder disorders: a systematic review of randomized clinical trials. *Br J Gen Pract* 1996; **46**:309–16.
20. Mink JH, Harris E, Rappaport M. Rotator cuff tears: evaluation using double-contrast shoulder arthrography. *Radiology* 1985;**157**:621–3.
21. Blanchard TK, Constant CR, Bearcroft PW, Marshall TJ, Dixon AK. Imaging of the rotator cuff: an arthrographic pitfall. *Eur Radiol* 1998; **8**:817–19.
22. Whiting P, Rutjes A, Dinnes J, Reitsma J, Bossuyt P, Kleijnen J. The development and validation of methods for assessing the quality and reporting of diagnostic studies. *Health Technol Assess* 2003; submitted.
23. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In Egger M, Smith GD, Altman DA, editors. *Systematic reviews in health care*. London: BMJ Books, 2001. pp. 248–82.
24. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III How to use an article about a diagnostic test A. Are the results of the study valid? *JAMA* 1994;**271**:389–91.

25. Smith-Bindman R, Hosmer W, Feldstein VA, Deeks JJ, Goldberg JD. Second-trimester ultrasound to detect fetuses with Down syndrome: a meta-analysis. *JAMA* 2001;**285**:1044–55.

26. Bearcroft PW, Blanchard TK, Dixon AK, Constant CR. An assessment of the effectiveness of magnetic resonance imaging of the shoulder: literature review. *Skeletal Radiol* 2000;**29**:673–9.

27. Oh CH, Schweitzer ME, Spettell CM. Internal derangements of the shoulder: decision tree and cost-effectiveness analysis of conventional arthrography, conventional MRI, and MR arthrography. *Skeletal Radiol* 1999;**28**:670–8.

28. Goupille P, Sibilia J. Local corticosteroid injections in the treatment of rotator cuff tendinitis (except for frozen shoulder and calcific tendinitis). Groupe Rhumatologique Français de l'Epaule (G.R.E.P.). *Clin Exp Rheumatol* 1996;**14**:561–6.

29. van der Windt DA, van der Heijden GJ, Scholten RJ, Koes BW, Bouter LM. The efficacy of non-steroidal anti-inflammatory drugs (NSAIDs) for shoulder complaints. A systematic review. *J Clin Epidemiol* 1995;**48**:691–704.

30. Calis M, Akgun K, Birtane M, Karacan I, Calis H, Tuzun F. Diagnostic values of clinical diagnostic tests in subacromial impingement syndrome. *Ann Rheum Dis* 2000;**59**:44–7.

31. Hertel R, Ballmer FT, Lombert SM, Gerber C. Lag signs in the diagnosis of rotator cuff rupture. *J Shoulder Elbow Surg* 1996;**5**:307–13.

32. Itoi E, Kido T, Sano A, Urayama M, Sato K. Which is more useful, the "full can test" or the "empty can test," in detecting the torn supraspinatus tendon? *Am J Sports Med* 1999; **27**:65–8.

33. Litaker D, Pioro M, El Bilbeisi H, Brems J. Returning to the bedside: using the history and physical examination to identify rotator cuff tears. *J Am Geriatr Soc* 2000;**48**:1633–7.

34. Lyons AR, Tomlinson JE. Clinical diagnosis of tears of the rotator cuff. *J Bone Joint Surg Br* 1992; **74**:414–15.

35. MacDonald PB, Clark P, Sutherland K. An analysis of the diagnostic accuracy of the Hawkins and Neer subacromial impingement signs. *J Shoulder Elbow Surg* 2000;**9**:299–301.

36. Read JW, Perko M. Shoulder ultrasound: diagnostic accuracy for impingement syndrome, rotator cuff tear, and biceps tendon pathology. *J Shoulder Elbow Surg* 1998;**7**:264–71.

37. Wnorowski DC, Levinsohn EM, Chamberlain BC, McAndrew DL. Magnetic resonance imaging assessment of the rotator cuff: is it really accurate? *Arthroscopy* 1997;**13**:710–19.

38. Wolf EM, Agrawal V. Transdeltoid palpation (the rent test) in the diagnosis of rotator cuff tears. *J Shoulder Elbow Surg* 2001;**10**:470–3.

39. Zaslav KR. Internal rotation resistance strength test: a new diagnostic test to differentiate intra-articular pathology from outlet (Neer) impingement syndrome in the shoulder. *J Shoulder Elbow Surg* 2001;**10**:23–7.

40. Crass JR, Craig EV, Feinberg SB. Ultrasonography of rotator cuff tears: a review of 500 diagnostic studies. *J Clin Ultrasound* 1988;**16**:313–27.

41. Middleton WD, Reinus WR, Totty WG, Melson CL, Murphy WA. Ultrasonographic evaluation of the rotator cuff and biceps tendon. *J Bone Joint Surg Am* 1986;**68**:440–50.

42. Wallny TA, Schild RL, Bertelsbeck DS, Hansmann ME, Kraft CN. Three-dimensional ultrasonography in the diagnosis of rotator cuff lesions. *Ultrasound Med Biol* 2001;**27**:745–9.

43. Gratz S, Koster G, Behr T, Vosshenrich R, Grabbe E, Becker W. Arthroscintigraphy in suspected rotator cuff rupture. *Nuklearmedizin* 1998;**37**:272–8.

44. Kurol M, Rahme H, Hilding S. Sonography for diagnosis of rotator cuff tear. Comparison with observations at surgery in 58 shoulders. *Acta Orthop Scand* 1991;**62**:465–7.

45. Olive RJ, Marsh HO. Ultrasonography of rotator cuff tears. *Clin Orthop* 1992;110–13.

46. Mack LA, Matsen FA, Kilcoyne RF, Davies PK, Sickler ME. US evaluation of the rotator cuff. *Radiology* 1985;**157**:205–9.

47. Mack LA, Gannon MK, Kilcoyne RF, Matsen RA. Sonographic evaluation of the rotator cuff. Accuracy in patients without prior surgery. *Clin Orthop* 1988;**234**:21–7.

48. Middleton WD, Edelstein G, Reinus WR, Melson GL, Totty WG, Murphy WA. Sonographic detection of rotator cuff tears. *Am J Roentgenol* 1985;**144**:349–53.

49. Bretzke CA, Crass JR, Craig EV, Feinberg SB. Ultrasonography of the rotator cuff. Normal and pathologic anatomy. *Invest Radiol* 1985;**20**:311–15.

50. Arslan G, Apaydin A, Kabaalioglu A, Sindel T, Luleci E. Sonographically detected subacromial/subdeltoid bursal effusion and biceps tendon sheath fluid: reliable signs of rotator cuff tear? *J Clin Ultrasound* 1999;**27**:335–9.

51. Brenneke SL, Morgan CJ. Evaluation of ultrasonography as a diagnostic technique in the assessment of rotator cuff tendon tears. *Am J Sports Med* 1992;**20**:287–9.

52. Farin PU, Jaroma H, Harju A, Soimakallio S. Shoulder impingement syndrome: sonographic evaluation. *Radiology* 1990;**176**:845–9.

53. Sonnabend DH, Hughes JS, Giuffre BM, Farrell R. The clinical role of shoulder ultrasound. *Aust NZ J Surg* 1997;**67**:630–3.

54. Takagishi K, Makino K, Takahira N, Ikeda T, Tsuruno K, Itoman M. Ultrasonography for diagnosis of rotator cuff tear. *Skeletal Radiol* 1996;**25**:221–4.

55. Teefey SA, Hasan SA, Middleton WD, Patel M, Wright RW, Yamaguchi K. Ultrasonography of the rotator cuff. A comparison of ultrasonographic and arthroscopic findings in one hundred consecutive cases. *J Bone Joint Surg Am* 2000; **82**:498–504.

56. Wiener SN, Seitz WH. Sonography of the shoulder in patients with tears of the rotator cuff: accuracy and value for selecting surgical options. *Am J Roentgenol* 1993;**160**:103–7.

57. De Muynck M. Sonography in the diagnosis of rotator cuff ruptures. *Eur J Phys Med Rehabil* 1994; **4**:130–3.

58. Misamore GW, Woodward C. Evaluation of degenerative lesions of the rotator cuff. A comparison of arthrography and ultrasonography. *J Bone Joint Surg Am* 1991;**73**:704–6.

59. Hodler J, Terrier B, von Schulthess GK, Fuchs WA. MRI and sonography of the shoulder. *Clin Radiol* 1991;**43**:323–7.

60. Naredo AE, Aguado P, Padron M, Bernad M, Uson J, Mayordomo L, *et al*. A comparative study of ultrasonography with magnetic resonance imaging in patients with painful shoulder. *J Clin Rheumatol* 1999;**5**:184–92.

61. Paavolainen P, Ahovuo J. Ultrasonography and arthrography in the diagnosis of tears of the rotator cuff. *J Bone Joint Surg Am* 1994;**76**:335–40.

62. Swen WA, Jacobs JW, Neve WC, Bal D, Bijlsma JW. Is sonography performed by the rheumatologist as useful as arthrography executed by the radiologist for the assessment of full thickness rotator cuff tears? *J Rheumatol* 1998;**25**:1800–6.

63. Swen WA, Jacobs JW, Algra PR, Manoliu RA, Rijkmans J, Willems WJ, *et al*. Sonography and magnetic resonance imaging equivalent for the assessment of full-thickness rotator cuff tears. *Arthritis Rheum* 1999;**42**:2231–8.

64. Nelson MC, Leather GP, Nirschl RP, Pettrone FA, Freedman MT. Evaluation of the painful shoulder. A prospective comparison of magnetic resonance imaging, computerized tomographic arthrography, ultrasonography, and operative findings. *J Bone Joint Surg Am* 1991;**73**:707–16.

65. Shipley ES. Ultrasound examination of the painful shoulder. *J Diagn Med Sonogr* 1985;**1**:261–7.

66. Martin-Hervas C, Romero J, Navas-Acien A, Reboiras JJ, Munuera L. Ultrasonographic and magnetic resonance images of rotator cuff lesions compared with arthroscopy or open surgery findings. *J Shoulder Elbow Surg* 2001;**10**:410–15.

67. van Moppes FI, Veldkamp O, Roorda J. Role of shoulder ultrasonography in the evaluation of the painful shoulder. *Eur J Radiol* 1995;**19**:142–6.

68. Drakeford MK, Quinn MJ, Simpson SL, Pettine KA. A comparative study of ultrasonography and arthrography in evaluation of the rotator cuff. *Clin Orthop* 1990;**253**:118–22.

69. Pattee GA, Snyder SJ. Sonographic evaluation of the rotator cuff: correlation with arthroscopy. *Arthroscopy* 1988;**4**:15–20.

70. Brandt TD, Cardone BW, Grant TH, Post M, Weiss CA. Rotator cuff sonography: a reassessment. *Radiology* 1989;**173**:323–7.

71. Burk DL, Karasick D, Kurtz AB, Mitchell DG, Rifkin MD, Miller CL, *et al*. Rotator cuff tears: prospective comparison of MR imaging with arthrography, sonography, and surgery. *Am J Roentgenol* 1989;**153**:87–92.

72. Miller CL, Karasick D, Kurtz AB, Fenlin JM. Limited sensitivity of ultrasound for the detection of rotator cuff tears. *Skeletal Radiol* 1989; **18**:179–83.

73. Ahovuo J, Paavolainen P, Bjorkenheim JM. Ultrasonography in lesions of the rotator cuff and biceps tendon. *Acta Radiol* 1989;**30**:253–5.

74. Vick CW, Bell SA. Rotator cuff tears: diagnosis with sonography. *Am J Roentgenol* 1990;**154**:121–3.

75. Soble MG, Kaye AD, Guay RC. Rotator cuff tear: clinical experience with sonographic detection. *Radiology* 1989;**173**:319–21.

76. Chiou HJ, Chou YH, Wu JJ, Hsu CC, Tiu CM, Chang CY. High-resolution ultrasonography of the musculoskeletal system: analysis of 369 cases. *J Med Ultrasound* 1999;**7**:212–18.

77. Roberts CS, Walker JA, Seligson D. Diagnostic capabilities of shoulder ultrasonography in the detection of complete and partial rotator cuff tears. *Am J Orthop* 2001;**30**:159–62.

78. van Holsbeeck MT, Kolowich PA, Eyler WR, Craig JG, Shirazi KK, Habra GK, *et al*. US depiction of partial-thickness tear of the rotator cuff. *Radiology* 1995;**197**:443–6.

79. Hodler J, Fretz CJ, Terrier F, Gerber C. Rotator cuff tears: correlation of sonographic and surgical findings. *Radiology* 1988;**169**:791–4.

80. Crass JR, Craig EV, Thompson RV, Feinberg SV. Ultrasonography of the rotator cuff: surgical correlation. *J Clin Ultrasound* 1984;**12**:487–92.

81. Irwig LM, Bossuyt PM, Glasziou PP, Gatsonis C, Lijmer JG. Designing studies to ensure that estimates of test accuracy will travel. In Knottnerus

JA, editor. The evidence base of clinical diagnosis. London: BMJ Books, 2002. pp. 95–116.

82. Torstensen ET, Hollinshead RM. Comparison of magnetic resonance imaging and arthroscopy in the evaluation of shoulder pathology. *J Shoulder Elbow Surg* 1999;**8**:42–5.

83. Balich SM, Sheley RC, Brown TR, Sauser DD, Quinn SF. MR imaging of the rotator cuff tendon: interobserver agreement and analysis of interpretive errors. *Radiology* 1997;**204**:191–4.

84. Quinn SF, Sheley RC, Demlow TA, Szumowski J. Rotator cuff tendon tears: evaluation with fat-suppressed MR imaging with arthroscopic correlation in 100 patients. *Radiology* 1995; **195**:497–500.

85. Robertson PL, Schweitzer ME, Mitchell DG, Schlesinger F, Epstein RE, Frieman BG, *et al*. Rotator cuff disorders: interobserver and intraobserver variation in diagnosis with MR imaging. *Radiology* 1995;**194**:831–5.

86. Reinus WR, Shady KL, Mirowitz SA, Totty WG. MR diagnosis of rotator cuff tears of the shoulder: value of using T2-weighted fat-saturated images. *Am J Roentgenol* 1995;**164**:1451–5.

87. Sahin-Akyar G, Miller TT, Staron RB, McCarthy DM, Feldman F. Gradient-echo versus fat-suppressed fast spin-echo MR imaging of rotator cuff tears. *Am J Roentgenol* 1998;**171**:223–7.

88. Birtane M, Calis M, Akgun K. The diagnostic value of magnetic resonance imaging in subacromial impingement syndrome. *Yonsei Med J* 2001;**42**:418–24.

89. Evancho AM, Stiles RG, Fajman WA, Flower SP, Macha T, Brunner MC, *et al*. MR imaging diagnosis of rotator cuff tears. *Am J Roentgenol* 1988;**151**:751–4.

90. Hodler J, Kursunoglu-Brahme S, Snyder SJ, Cervilla V, Karzel RP, Schweitzer ME, *et al*. Rotator cuff disease: assessment with MR arthrography versus standard MR imaging in 36 patients with arthroscopic confirmation. *Radiology* 1992; **182**:431–6.

91. Jaovisidha S, Jacobson JA, Lenchik L, Resnick D. MR imaging of rotator cuff tears: is there a diagnostic benefit of shoulder exercise prior to imaging. *Clin Imaging* 1999;**23**:249–53.

92. Patten RM, Spear RP, Richardson ML. Diagnostic performance of magnetic resonance imaging for the diagnosis of rotator cuff tears using supplemental images in the oblique sagittal plane. *Invest Radiol* 1994;**29**:87–93.

93. Tuite MJ, Asinger D, Orwin JF. Angled oblique sagittal MR imaging of rotator cuff tears: comparison with standard oblique sagittal images. *Skeletal Radiol* 2001;**30**:262–9.

94. Morrison DS, Ofstein R. The use of magnetic resonance imaging in the diagnosis of rotator cuff tears. *Orthopedics* 1990;**13**:633–7.

95. Kneeland JB, Middleton WD, Carrera GF, Zeuge RC, Jesmanowicz A, Froncisz W, *et al*. MR imaging of the shoulder: diagnosis of rotator cuff tears. *Am J Roentgenol* 1987;**149**:333–7.

96. Wang YM, Shih TT, Jiang CC, Su CT, Huang KM, Hang YS, *et al*. Magnetic resonance imaging of rotator cuff lesions. *J Formos Med Assoc* 1994; **93**:234–9.

97. Yagci B, Manisali M, Yilmaz E, Ozkan M, Ekin A, Ozaksoy D, *et al*. Indirect MR arthrography of the shoulder in detection of rotator cuff ruptures. *Eur Radiol* 2001;**11**:258–62.

98. Blanchard TK, Bearcroft PW, Constant CR, Griffin DR, Dixon AK. Diagnostic and therapeutic impact of MRI and arthrography in the investigation of full-thickness rotator cuff tears. *Eur Radiol* 1999; **9**:638–42.

99. Tuite MJ, Yandow DR, De Smet AA, Orwin JF, Quintana FA. Effect of field of view on MR diagnosis of rotator cuff tears. *Skeletal Radiol* 1995;**24**:495–8.

100. Pfirrmann CW, Zanetti M, Weishaupt D, Gerber C, Hodler J. Subscapularis tendon tears: detection and grading at MR arthrography. *Radiology* 1999; **213**:709–14.

101. Funke M, Kopka L, Vosshenrich R, Oestmann JW, Grabbe E. MR arthrography in the diagnosis of rotator cuff tears. Standard spin-echo alone or with fat suppression? *Acta Radiol* 1996;**37**:627–32.

102. Loew R, Kreitner KF, Runkel M, Zoellner J, Thelen M. MR arthrography of the shoulder: comparison of low-field (0.2 T) vs high-field (1.5 T) imaging. *Eur Radiol* 2000;**10**:989–96.

103. Binkert CA, Zanetti M, Gerber C, Hodler J. MR arthrography of the glenohumeral joint: two concentrations of gadoteridol versus Ringer solution as the intraarticular contrast material. *Radiology* 2001;**220**:219–24.

104. Blanchard TK, Mackenzie R, Bearcroft PW, Sinnatamby R, Gray A, Lomas DJ, *et al*. Magnetic resonance imaging of the shoulder: assessment of effectiveness. *Clin Radiol* 1997;**52**:363–8.

105. Sher JS, Iannotti JP, Williams GR, Herzog RJ, Kneeland JB, Lisser S, *et al*. The effect of shoulder magnetic resonance imaging on clinical decision making. *J Shoulder Elbow Surg* 1998;**7**:205–9.

106. Blanchard TK, Bearcroft PW, Maibaum A, Hazelman BL, Sharma S, Dixon AK. Magnetic resonance imaging or arthrography for shoulder problems: a randomised study. *Eur J Radiol* 1999; **30**:5–10.

107. Zanetti M, Jost B, Lustenberger A, Hodler J. Clinical impact of MR arthrography of the shoulder. *Acta Radiol* 1999;**40**:296–302.

108. Blanchard TK, Bearcroft PW, Dixon AK, Lomas DJ, Teale A, Constant CR, *et al*. Magnetic resonance imaging or arthrography of the shoulder: which do patients prefer? *Br J Radiol* 1997;**70**:786–90.

109. Spielberger CD, Gorsuch RL and Lushene RE. The state-trait anxiety inventory (STAI) – test manual for form X. Palo Alto, CA: California Psychological Press, 1968.

110. Song F, Eastwood A, Gilbody S, Duley D, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;**4**(10).

111. Neer CS. Impingement lesions. *Clin Orthop* 1983; **173**:70–7.

112. Yergason RM. Supination sign. *J Bone Joint Surg Am* 1931;**13**:160.

113. Kieft GJ, Bloem JL, Rozing PM, Obermann WR. Rotator cuff impingement syndrome: MR imaging. *Radiology* 1988;**166**:211–14.

114. Traughber PD, Goodwin TE. Shoulder MRI: arthroscopic correlation with emphasis on partial tears. *J Comput Assist Tomogr* 1992;**16**:129–33.

115. Tuite MJ, Yandow DR, DeSmet AA, Orwin JF, Quintana FA. Diagnosis of partial and complete rotator cuff tears using combined gradient echo and spin echo imaging. *Skeletal Radiol* 1994; **23**:541–5.

116. Kelly BT, Kadrmas WR, and Speer KP. The manual muscle examination for rotator cuff strength: an electromyographic investigation. *Am J Sports Med* 1996;**24**:581–8.

117. Jobe FW, Moynes DR. Delineation of diagnostic criteria and a rehabilitation program for rotator cuff injuries. *Am J Sports Med* 1982; **10**:336–9.

118. Hawkins RH, Kennedy JC. Impingement syndrome in athletes. *Am J Sports Med* 1980; **8**:151–8.

119. Codman EA. Rupture of the supraspinatus tendon. 1911. *Clin Orthop* 1990;**254**:3–26.

120. McLaughlin HL. Rupture of the rotator cuff. *J Bone Joint Surg Am* 1962;**44**:979–93.

121. Middleton WD, Edelstein G, Reinus WR, Melson CL, Totty WG, Murphy WA. Sonographic detection of rotator cuff tears. *Am J Ultrasound* 1985; **20**:311.

122. Middleton WD, Edelstein G, Reinus WR et al. Ultrasonography of the rotator cuff: technique and normal anatomy. *J Ultrasound Med* 1984;**3**(12):549–51.

123. Mack LA, Nyberg DA, Matsen FA. Sonographic evaluation of the rotator cuff. *Radiol Clin North Am* 1988;**26**:161–77.

124. Crass JR, Craig EV, Feinberg SB. Sonography of the post-operative rotator cuff. *AJR* 1986:**146**:561–64.

125. Crass JR, van de Vetge GL, Harkavy LA. Tendon echogenicity: *ex vivo* study. *Radiology* 1988;**167**:499–501.

126. Chiou HJ, Hsu CC, Chou YH, Tiu CM, Jim YF, Wu JJ, Chang CY. Sonographic signs of complete rotator cuff tears. *Chung Hua I Hsueh Tsa Chih (Taipei)* 1996;**58**(6):428–34.

127. Crass JR, Craig EV, Feinberg SB. The hyperextended internal rotation view in rotator cuff ultrasonography. *J Clin Ultrasound* 1987;**15**:416–20.

128. Katthagen BA, Ludwig FJ, Dieudonne M. Schultersonographie. Technik – Anatomie – Pathologie, 1988; Stuttgart:Thieme.

129. Crass JR, Craig EV, Feinberg SB. Clinical significance of sonographic findings in the abnormal but intact rotator cuff: a preliminary report. *J Clin Ultrasound* 1988;**16**: 625–34.

130. Middleton WD, Reinus WR, Melson GL, Totty WG, Murphy WA. Pitfalls of rotator cuff sonography. *AJR Am J Roentgenol.* 1986;**146**: 555–60.

131. Mack LA, Gannon MK, Kilcoyne RF, Matsen RA. Sonographic evaluation of the rotator cuff. Accuracy in patients without prior surgery. *Clin Orthop* 1988;**234**:21–7.

132. Gristina AG. Personal Communication 1987. cited in Drakeford 1990.

133. Crass JR, Craig EV, Bretzke CA, Feinberg SB. Ultrasonography of the rotator cuff. *Radiographics* 1985;**5**:941–53.

134. Rapf C, Furtschegger A, Resch H. Die sonographie als neues diagnostisches verfahren zur abklarung von schulterbeschwerden. *Fortschr Roentgenstr* 1986;**3**(288):295.

135. Collins RA, Gristina AG, Carter RE, Webb LX, Voytek A. Ultrasonography of the shoulder: Static and dynamic imaging. *Orthop Clin North Am* 1987;**8**:351.

136. Middleton WD. Ultrasonography of the shoulder. *Radiol Clin North Am* 1992;**30**,927–40.

137. Farin PU, Jaroma H, Harju A, Soimakallio S. Medical displacement of the biceps brachii tendon: evaluation with dynamic sonography during maximal external shoulder rotation. *Radiology* 1995;**195**:845–8.

138. Van Holsbeeck MT, Introcaso JH. Sonography of the shoulder. In Van Holsbeeck MT, Introcaso JH, editors. Musculoskeletal ultrasound. Chicago: Mosby Year Book, 1991; pp. 265–84.

139. Van Holsbeeck MT, Introcaso JH. Musculoskeletal ultrasonography. *Radiol Clin North Am* 1992;**30**:907–25.

140. Hedtman A, Fett H. Atlas und Lehrbuch der Schultersonographie. Enke:Verlag, 1988.

141. Bachmann G, Melzer C, Heinrichs CM, Mohring B, Rominger MB. Diagnosis of rotator cuff lesions: comparison of US and MRI on 38 joint specimens. *Eur Radiol* 1997;**7**:192–7.

142. Hedtman A, Fett H. Ultrasonography of the shoulder in subacromial syndromes with disorders and injuries of the rotator cuff. *Orthopaede* 1995;**24**:498–508.

143. Fukuda H, Hamada N, Nakajima T. Partial thickness tears of the rotator cuff. *Int Orthop* 1996;**20**:257–65.

144. Crass JR, Craig EV. Non-invasive imaging of the rotator cuff. *Orthopedics* 1988;**11**:57–64.

145. Rafii M, Firooznia H, Sherman O, Minkoff J, Weinreb J, Golimbu C, Gidumal R, Schinella R, Zaslav K. Rotator cuff lesions: signal patterns at MR imaging. *Radiology* 1990;**177**(3): 817–23.

146. Holt RG, Helms CA, Steinbach L, Neumann C, Munk PL, Genant HK. Magnetic resonance imaging of the shoulder: rationale and current applications. *Skeletal Radiol* 1990;**19**: 5–14.

147. Zlatkin MB, Reicher MA, Kellerhouse LE, Resnick D. The painful shoulder: MR imaging of the glenohumeral joint. *J Comput Assist Tomogr* 1998;**12**:995–1001.

Appendix 1

Search strategy

Diagnosis terms

Clinical near4 assess*
Diagnos*
Diagnostic imaging – MeSH explode
Tomography
Ultrasonography – Mesh and ft
ultrasound
computed tomography arthrography (CTA)
computed tomography or CT
computerized tomography or CT scan
magnetic resonance imaging or MRI – explode or ft
magnetic resonance near2 arthrography or MRA or MR
sonograp*
MR imaging
arthroscopy
electromyography or EMG
electromyogram
arthrogram
imaging techniques
imaging modality
x-ray

Shoulder terms

Shoulder* near pain*

Causes of shoulder pain:

chronic shoulder impingement syndrome
disorders of impingement
rotator cuff tears
biceps tenosynovitis
instability

bursitis
rotator cuff
tendonitis
bicipital tendinitis
shoulder near arthritis
labral tear* and shoulder*
shoulder capsulitis
shoulder joint
frozen shoulder
shoulder disorder
shoulder injuri*
joint instability
shoulder dislocation*
tendinitis

Appendix 2

Quality assessment tool

Adapted from the QUADAS tool, developed by University of York.[22]

1. Spectrum

The spectrum of patients refers not only to the severity of the underlying target condition, but also to demographic features and to the presence of differential diagnosis and/or comorbidity. It is therefore important that diagnostic test evaluations include an appropriate spectrum of patients for the test under investigation and that a clear definition of the characteristics of the included patients is provided. The judgement should be based on both the method of recruitment and the characteristics of those recruited. Reviewers should pre-specify in the protocol of the review what spectrum of patients would be acceptable taking factors such as disease prevalence and severity, age and sex, into account.

In this review a judgement on the spectrum of included patients was made on the basis of:

- whether the study was prospective or retrospective
- whether patients were recruited consecutively
- if they were only included if they had been selected to undergo the reference test
- the study setting
- the age, sex and pain duration of included patients.

In studies evaluating clinical examination, the aim is to identify those people who would most benefit from further diagnostic investigation or indeed therapy. It was therefore important that studies evaluating clinical examination are conducted on a relatively unselected population. Only where sufficient assurance that this was in fact the case did studies score 'yes' on this item.

For the studies evaluating imaging tests, almost all included only 'problem' patients or those who had already failed conservative treatment. As imaging tests may be used either in a specialist or more generalist setting, this did not automatically mean that these studies scored 'no' on this item.

2. Eligibility criteria

Whether studies have provided a clear definition of the criteria used as selection criteria for entry into the study. If all relevant information regarding how participants were selected for inclusion in the study was provided, this item was scored as 'yes'. If study selection criteria were not clearly reported, for example, 'patients with shoulder pain' or 'patients referred for ultrasound' or for arthrography, then this item was scored as 'no'. Where selection criteria were only partially reported and insufficient information was given to score this item as 'yes', for example, 'physical examination suggestive of rotator cuff injury', then it was scored as 'unclear'.

3. Appropriate reference test

If the reference standard was likely to correctly classify the target condition, then this item was scored 'yes'. Arthroscopy, arthrography and surgery used either alone or in combination were all considered to be appropriate reference tests, despite disagreements in the literature as to which of these is the best, especially regarding arthroscopy versus arthrography. Studies using other tests such as MRI or the subacromial injection test (SIT) as the reference test were scored as unclear as the true accuracy of these tests have yet to be established.

4. Disease progression bias

Disease progression bias occurs where there is a delay between performance of the index and reference tests such that misclassification due to spontaneous recovery or advancement of disease may occur. This was judged to be present where the mean/median interval between tests was reported and was greater than 1 month. This item was scored as 'no' when the index and reference tests were performed within 1 month of each other, and as 'unclear' if the interval between tests was not clearly reported.

5. Partial verification

If it is clear from the study that all patients who received the index test went on to receive verification of their disease status using a reference standard, even if this reference standard was not the same for all patients, then this item was scored as 'yes'. If some of the patients who received the index test did not receive verification of their true disease state, then this item was scored as 'no'.

6. Differential verification

Differential verification bias occurs when some of the index test results are verified by a different reference standard. This is especially a problem if these reference standards differ in their definition of the target condition. Differential verification bias usually occurs when patients testing positive on the index test receive a more accurate, often invasive, reference standard than those with negative test results. If it is clear that all patients received verification of their true disease status using the same reference standard, then this item was scored as 'yes'. If some patients received verification using a different reference standard (regardless of whether or not the decision was based on the index test result), this item was scored as 'no'.

7. Incorporation bias

When the result of the index test is used in establishing the final diagnosis, incorporation bias may occur.

8a and b. Sufficient details of index and reference tests provided

If the study reports sufficient details to permit replication of the index test and reference standard, then these items should be scored as 'yes'.

9a and b. Test review and diagnostic review bias

Interpretation of the results of the index test may be influenced by knowledge of the results of the reference standard, and vice versa. This is known as review bias, and may lead to inflated measures of diagnostic accuracy. The extent to which this will affect test results is related to the degree of subjectivity involved in the interpretation of each test result. The more subjective the interpretation, the more likely it is that the interpreter can be influenced by the results of the index test in interpreting the reference standard, and vice versa. The interpretation of clinical examination or any of the imaging tests considered here may be highly subjective and could be considerably influenced by knowledge of the reference test result. Although in all studies the index test would have been performed first, the results of the index test may still have been interpreted in the light of the results of the reference standard, especially where test results were identified from retrospective chart review.

If the study clearly states that the test results (index or reference standard) were interpreted blind to the results of the other test, then these items were scored as 'yes'. If this did not appear to be the case, they were scored as 'no'. If this information was not reported by the study, then it was scored as 'unclear'.

10. Appropriate clinical information available

The availability of information on clinical data during interpretation of test results may affect estimates of test performance. Clinical data are here defined broadly to include any information relating to the patient obtained by direct observation such as age, sex and symptoms. The knowledge of such factors can influence the diagnostic test result if the test involves an interpretative component. If clinical data will be available when the test is interpreted in practice, then this should also be available when the test is evaluated. If, however, the index test is intended to replace other clinical tests, then clinical data should not be available. It is therefore important to determine what information will be available when test results are interpreted in practice before assessing studies for this item.

As clinical data would normally be available when clinical examination is conducted or when imaging tests are interpreted in practice, then similar data should be available when interpreting the index test in the study. If such data were not available, then this item was scored as 'no'. If this information is not reported by the study, then it was scored as 'unclear'.

11. Uninterpretable/intermediate test results reported

A diagnostic test can produce an uninterpretable/indeterminate/intermediate result with varying frequency depending on the test. These problems are often not reported in diagnostic accuracy studies with the uninterpretable results simply removed from the analysis. This may lead to the biased assessment of the test characteristics. Whether bias will arise depends on the possible correlation between uninterpretable test results and the true disease status. If uninterpretable results occur randomly and are not related to the true disease status of the individual, then, in theory, these should not have any effect on test performance. Whatever the cause of uninterpretable results, it is important that these are reported so that the impact of these results on test performance can be determined.

If it is clear that all test results, including uninterpretable/indeterminate/intermediate, were reported, then this item was scored as 'yes'. If it appeared that such results occurred but were not reported, this item was scored as 'no'. If it was not clear whether all study results were reported, this item was scored as 'unclear'.

12. Withdrawal bias

This occurs when patients withdraw from the study before the results of both the index test and reference standard are known. If patients lost to follow-up differ systematically from those who remain, for whatever reason, then estimates of test performance may be biased.

If it was clear what happened to all patients who entered the study, for example if a flow diagram of study participants was reported, or reasons for exclusion of patients were provided, then this item was scored as 'no'. If it appeared that some of the participants who entered the study did not complete the study, that is, did not receive both the index test and reference standard, and these patients were not accounted for, then this item was scored as 'yes'. If it was not clear whether all patients who entered the study were accounted for, then this item was scored as 'unclear'.

Appendix 3

Data extraction form

First Author | Public Year: 0 | Ref ID: 1

Review Inclusion Criteria Met:

☐ Test ☐ Reference test ☐ Population Study Type:

Study details: No. of index tests evaluated:

Index Test 1: Public status:

Index Test 2:

Index Test 3: Country:

ref Test 1: Study years:

Population:

Study Design: Study Setting:

Prosp/Retrosp: Study Setting 2

Sample Size:

Sample Index: Sample Ref:

Study eligibility criteria

1

2

3

4

Patient characteristics:

Age Previous shoulder pain:

Sex: Comorbidities:

Race: Pain duration:

Other:

Further test details:

Test 1 Details:

Test 1 Categories:

Test 2 Details:

Test 2 Categories:

Test 3 Details:

Test 3 Categories:

Indeterminate results:

Ref Details:

Ref Categories:

Interval between tests

Further investigations

Test interpreter

Qualifications:

Number:

Ref test interprete

Quality Assessment:

Patient selection biases

Referral bias

Study establishment

Referral establishment

Access to establishment

Patient filtering bias

Eligibility criteria stated

Co-intervention bias present

Co-intervention bias avoided

Patient cohort bias

Clinical details given

Cornorbid details given

Other

Appropriate spectrum

Patient recruitment

Study design

Test selection and execution

Test details

Details of index test

Details of reference test

Appropriate reference test

Normal defined (index test)

Application of gold standard

Verification bias present

Work-up bias present

Differential VB present

Incorporation bias present

Reference test enhancement

Independence of test interpretation

Test review bias present

Diagnostic review bias present

Clinical review bias present

Prospective review of images

Observer variability

Single observer of index test?

Separate results per observer?

Inter-observer variability

Consensus decision?

Intra-observer variability

Measurement of results

Disease progression bias

Withdrawal bias present

Indeterminate results missing

Loss to follow-up

Cost data:

Study Comment

Study results:

☐ Data for 2×2 table provided ☐ If not, can data for 2×2 be estimated

DETECTION OF ANY ROTATOR CUFF TEAR

Author: Year: Ref ID: 1

Comparison 1

vs.

Cell a: 0 Cell b: 0

Cell c: 0 Cell d: 0

Reported sensitivity: Revised sensitivity:

Reported specificity: Revised specificity:

Reported PPV: Revised PPV:

Reported NPV: Revised NPV:

Any other reported state:

Comparison 2

vs.

Cell a: 0 Cell b: 0

Cell c: 0 Cell d: 0

Reported sensitivity: Revised sensitivity:

Reported specificity: Revised specificity:

Reported PPV: Revised PPV:

Reported NPV: Revised NPV:

Any other reported state:

Comparison 3

vs.

Cell a: 0 Cell b: 0

Cell c: 0 Cell d: 0

Reported sensitivity: Revised sensitivity:

Reported specificity: Revised specificity:

Reported PPV: Revised PPV:

Reported NPV: Revised NPV:

Comparison 3 Other

Appendix 4

Reasons for exclusion

Adolfsson L, Lysholm J. Arthroscopy for the diagnosis of shoulder pain. *Int Orthop* 1991;**15**:275–8.
Reason for exclusion: wrong index test – arthroscopy.

Ahovuo J, Paavolainen P, Slatis P. Diagnostic value of sonography in lesions of the biceps tendon. *Clin Orthop* 1986;184–8.
Reason for exclusion: study type – not accuracy.

Allmann KH, Schafer O, Hauer M, Winterer J, Laubenberger J, Reichelt A, *et al*. Indirect MR arthrography of the unexercised glenohumeral joint in patients with rotator cuff tears. *Invest Radiol* 1999;**34**:435–40.
Reason for exclusion: wrong population – RCTs

Beltran J, Gray LA, Bools JC, Zuelzer W, Weis LD, Unverferth LJ. Rotator cuff lesions of the shoulder: evaluation by direct sagittal CT arthrography. *Radiology* 1986;**160**:161–5.
Reason for exclusion: ineligible index test – CT.

Ben Yishay AF, Zuckerman JD, Gallagher MF, Cuomo F, Speer KP, Hannafin JA, *et al*. Pain inhibition of shoulder strength in patients with impingement syndrome. An evaluation of the shoulder relocation test. *Orthopedics* 1994;**17**:685.
Reason for exclusion: no reference test.

Bencardino JT, Beltran J, Rosenberg ZS, Rokito A, Schmahmann S, Mota J, *et al*. Superior labrum anterior–posterior lesions: diagnosis with MR arthrography of the shoulder. *Radiology* 2000; **214**:267–71.
Reason for exclusion: wrong population – labral lesions.

Bennett WF. Specificity of the Speed's test: arthroscopic technique for evaluating the biceps tendon at the level of the bicipital groove. *Arthroscopy* 1998; **14**:789–96.
Reason for exclusion: wrong population – labral tears.

Berg EE, Ciullo JV. A clinical test for superior glenoid labral or 'SLAP' lesions. *Clin J Sport Med* 1998; **8**:121–3.
Reason for exclusion: wrong population – labral lesions.

Bjorksten MG, Boquist B, Talback M, Edling C. The validity of reported musculoskeletal problems. A study of questionnaire answers in relation to diagnosed disorders and perception of pain. *Appl Ergon* 1999; **30**:325–30.
Reason for exclusion: not investigating accuracy.

Blanchard TK, Constant CR, Bearcroft PW, Marshall TJ, Dixon AK. Imaging of the rotator cuff: an arthrographic pitfall. *Eur Radiol* 1998;**8**:817–19.
Reason for exclusion: case study.

Bretzke CA, Crass JR, Craig EV, Feinberg SB. Ultrasonography of the rotator cuff. Normal and pathologic anatomy. *Invest Radiol* 1985;**20**:311–15.
Reason for exclusion: reference test – limited to index test positives.

Bruyn GA, Rijnks J, den Hollander H, Griep EN, Leeuwarden J. Comparison of sonography and arthroscopy in patients with shoulder complaints. *Arthritis Rheum* 1999;**42**:S270.
Reason for exclusion: abstract only.

Buirski G. Magnetic resonance imaging in acute and chronic rotator cuff tears. *Skeletal Radiol* 1990; **19**:109–11.
Reason for exclusion: study type – not accuracy.

Burk DL, Karasick D, Mitchell DG, Rifkin MD. MR imaging of the shoulder: correlation with plain radiography. *Am J Roentgenol* 1990;**154**:549–53.
Reason for exclusion: study type – not accuracy.

Callaghan JJ, McNiesh LM, DeHaven JP, Savory CG, Polly DW. A prospective comparison study of double contrast computed tomography (CT) arthrography and arthroscopy of the shoulder. *Am J Sports Med* 1988;**16**:13–20.
Reason for exclusion: ineligible index test – CTA.

Cartland JP, Crues JV, Stauffer A, Nottage W, Ryu RK. MR imaging in the evaluation of SLAP injuries of the shoulder: findings in 10 patients. *Am J Roentgenol* 1992;**159**:787–92.
Reason for exclusion: wrong population – labral lesions.

Chan KM, Hsu SYC. Diagnostic arthroscopy of the shoulder – a prospective review of 115 cases with clinical and arthroscopic correlation. *J West Pac Orthop Assoc* 1991;**28**:39–47.
Reason for exclusion: reference test – limited to index test positives.

Chan WP, Jaw W, Mao C. Frozen shoulder – comparison of shoulder arthrography and MR- imaging for assessment of coexisting clinically occult rotator cuff tears. *Radiology* 1995;**197**:227.
Reason for exclusion: wrong population – adhesive capsulitis.

Chandnani VP, Yeager TD, DeBerardino T, Christensen K, Gagliardi JA, Heitz DR, *et al*. Glenoid labral tears: prospective evaluation with MRI imaging, MR arthrography, and CT arthrography. *Am J Roentgenol* 1993;**161**:1229–35.
Reason for exclusion: wrong population – labral tears.

Chandnani VP, Gagliardi JA, Murnane TG, Bradley YC, DeBerardino TA, Spaeth J, *et al*. Glenohumeral ligaments and shoulder capsular mechanism: evaluation with MR arthrography. *Radiology* 1995; **196**:27–32.
Reason for exclusion: wrong population – shoulder instability.

Chen JD, Jim YF, Chang CY. MR imaging of rotator cuff impingement: correlation with full-thickness rotator cuff tear. *Chung Hua I Hsueh Tsa Chih (Taipei)* 1996; **58**:198–204.
Reason for exclusion: study type – not accuracy.

Cook KF, Gartsman GM, Roddey TS, Olson SL. The measurement level and trait-specific reliability of 4 scales of shoulder functioning: An empiric investigation. *Arch Phys Med Rehabil* 2001; **82**:1558–65.
Reason for exclusion: study type – not accuracy.

Cvitanic O, Tirman PF, Feller JF, Bost FW, Minter J, Carroll KW. Using abduction and external rotation of the shoulder to increase the sensitivity of MR arthrography in revealing tears of the anterior glenoid labrum. *Am J Roentgenol* 1997;**169**:837–44.
Reason for exclusion: wrong population – labral tears.

D'Alessandro DF, Fleischli JE, Connor PM. Superior labral lesions: diagnosis and management. *J Athletic Training* 2000;**35**:286–92.
Reason for exclusion: not primary study; wrong population – labral tears.

de Winter AF, Jans MP, Scholten RJ, Deville W, van Schaardenburg D, Bouter LM. Diagnostic classification of shoulder disorders: interobserver agreement and determinants of disagreement. *Ann Rheum Dis* 1999;**58**:272–7.
Reason for exclusion: not diagnostic accuracy.

Emig EW, Schweitzer ME, Karasick D, Lubowitz J. Adhesive capsulitis of the shoulder: MR diagnosis. *Am J Roentgenol* 1995;**164**:1457–9.
Reason for exclusion: wrong population – adhesive capsulitis.

Farin P, Jaroma H. Sonographic detection of tears of the anterior portion of the rotator cuff (subscapularis tendon tears). *J Ultrasound Med* 1996;**15**:221–5.
Reason for exclusion: no specificity data.

Farin PU, Kaukanen E, Jaroma H, Vaatainen U, Miettinen H, Soimakallio S. Site and size of rotator-cuff tear. Findings at ultrasound, double-contrast arthrography, and computed tomography arthrography with surgical correlation. *Invest Radiol* 1996;**31**:387–94.
Reason for exclusion: ineligible index test – CTA.

Farley TE, Neumann CH, Steinbach LS, Jahnke AJ, Petersen SS. Full-thickness tears of the rotator cuff of the shoulder: diagnosis with MR imaging. *Am J Roentgenol* 1992;**158**:347–51.
Reason for exclusion: insufficient data.

Fermand M, Blanchard JP, Vergeron H, Goldberg D. Rotator cuff imaging using bursography coupled to helical computed arthrotomography. *Rev Rhum Engl Ed* 1999;**66**:131–5.
Reason for exclusion: reference test – CTA.

Fermand M, Hassen CS, Ariche L, Samuel P , Postel JM, Blanchard JP, *et al*. Ultrasound investigation of the rotator cuff after computed arthrotomography coupled to bursography. *Joint Bone Spine* 2000; **67**:310–4.
Reason for exclusion: wrong reference test – CTA.

Finkel LI, Berg DJ, Davis JL. Double-contrast computed tomography of the shoulder. *J Am Osteopath Assoc* 1989;**89**:1017–20.
Reason for exclusion: ineligible index test – CTA.

Flannigan B, Kursunoglu-Brahme S, Snyder S, Karzel R, Del Pizzo W, Resnick D. MR arthrography of the shoulder: comparison with conventional MR imaging. *Am J Roentgenol* 1990;**155**:829–32.
Reason for exclusion: insufficient data.

Guckel C, Nidecker A. MR arthrographic findings in tenosynovitis of the long bicipital tendon of the shoulder. *Skeletal Radiol* 1998;**27**:7–12.
Reason for exclusion: reference test – limited to index test positives.

Gusmer PB, Potter HG, Schatz JA, Wickiewicz TL, Altchek DW, O'Brien SJ, *et al*. Labral injuries: accuracy of detection with unenhanced MR imaging of the shoulder. *Radiology* 1996;**200**:519–24.
Reason for exclusion: wrong population – labral tears.

Habibian A, Stauffer A, Resnick D, Reicher MA, Rafii M, Kellerhouse L, *et al*. Comparison of conventional and computed arthrotomography with MR imaging in the evaluation of the shoulder. *J Comput Assist Tomogr* 1989;**13**:968–75.
Reason for exclusion: wrong reference test – CTA.

Hame SL, Nuccion S, Chuan J, Seeger L. Accuracy of MRI and MRA in the diagnosis of superior labrum anterior and posterior (SLAP) lesions. *Med Sci Sports Exerc* 2001;**33**:S275.
Reason for exclusion: abstract only; wrong population – labral lesions.

Hawkins RH, Kennedy JC. Impingement syndrome in athletes. *Am J Sports Med* 1980;**8**:151–8.
Reason for exclusion: not primary study.

Hodler J, Kursunoglu-Brahme S, Flannigan B, Snyder SJ, Karzel RP, Resnick D. Injuries of the superior portion of the glenoid labrum involving the insertion of the biceps tendon: MR imaging findings in nine cases. *Am J Roentgenol* 1992;**159**:565–8.
Reason for exclusion: wrong population – labral tears.

Horii M, Takubo Y, Yamaguchi J, Kurokawa M, Kubo T, Hirasawa Y. The diagnostic usefulness of magnetic resonance imaging for a partial-thickness rotator cuff tear. *J Orthop Surg* 1998;**6**:53–8.
Reason for exclusion: reference test – limited to index test positives.

Hutton J. Assessing the cost-effectiveness of imaging technology: recent results, problems and progress in the UK. *Eur Radiol* 2000;**10**:S427–9.
Reason for exclusion: review article.

Huylebroek J, van Hedent E, van Overschelde J. Correlation of computed arthrotomography with arthroscopy of the glenohumeral joint. *Acta Orthop Belg* 1991;**57**:S183–8.
Reason for exclusion: ineligible index test – CTA.

Imhoff AB, Hodler J. Correlation of MR imaging, CT arthrography, and arthroscopy of the shoulder. *Bull Hosp Jt Dis* 1996;**54**:146–52.
Reason for exclusion: insufficient data.

Jahnke AH, Petersen SA, Neumann C, Steinbach L, Morgan F. A prospective comparison of computerized arthrotomography and magnetic resonance imaging of the glenohumeral joint. *Am J Sports Med* 1992; **20**:695–700.
Reason for exclusion: wrong population – labral tears.

Jason M. Magnetic resonance imaging MRI scanning as a diagnostic tool in the evaluation of shoulder impingement. *Arthritis Rheum* 1988;**31**:S101.
Reason for exclusion: insufficient data.

Jee WH, McCauley TR, Katz LD, Matheny JM, Ruwe PA, Daigneault JP. Superior labral anterior posterior (SLAP) lesions of the glenoid labrum: reliability and accuracy of MR arthrography for diagnosis. *Radiology* 2001;**218**:127–32.
Reason for exclusion: wrong population – labral lesions.

Jobe FW, Jobe CM. Painful athletic injuries of the shoulder. *Clin Orthop* 1983;**173**:117–24.
Reason for exclusion: not assessing diagnostic accuracy.

Kalke SJ, Perera SD, Dasgupta B. Ultrasonography in the evaluation of shoulder pain. *Arthritis Rheum* 1999;**9**:S150.
Reason for exclusion: insufficient data; no gold standard.

Kibler WB. Specificity and sensitivity of the anterior slide test in throwing athletes with superior glenoid labral tears. *Arthroscopy* 1995;**11**:296–300.
Reason for exclusion: wrong population – labral tears.

Kim SH, Ha KI, Ahn JH, Kim SH, Choi HJ. Biceps load test II: a clinical test for SLAP lesions of the shoulder. *Arthroscopy* 2001;**17**:160–4.
Reason for exclusion: wrong population – labral lesions.

Kneisl JS, Sweeney HJ, Paige ML. Correlation of pathology observed in double contrast arthrotomography and arthroscopy of the shoulder. *Arthroscopy* 1988;**4**:21–4.
Reason for exclusion: wrong index test – arthroscopy and double contrast arthrotomography.

Kyung NR, Sun WL, Yong GR, Jae HL. Adhesive capsulitis of the shoulder joint: usefulness of dynamic sonography. *J Ultrasound Med* 1993;**12**:445–9.
Reason for exclusion: wrong population – adhesive capsulitis.

Legan JM, Burkhard TK, Goff WB, Balsara ZN, Martinez AJ, Burks DD, *et al*. Tears of the glenoid labrum: MR imaging of 88 arthroscopically confirmed cases. *Radiology* 1991;**179**:241–6.
Reason for exclusion: wrong population – labral tears.

Leroux JL, Thomas E, Bonnel F, Blotman F. Diagnostic value of clinical tests for shoulder impingement syndrome. *Rev Rhum Engl Ed* 1995; **62**:423–8.
Reason for exclusion: all had impingement syndrome – no specificity data.

Li XX, Schweitzer ME, Bifano JA, Lerman J, Manton GL, El Noueam KI. MR evaluation of subscapularis tears. *J Comput Assist Tomogr* 1999;**23**:713–17.
Reason for exclusion: reference test – limited to index test positives.

Litaker D, Pioro MH, El Bilbeise H, Brems JJ, Cash JM. Does the medical history and physical examination accurately predict the presence or absence of rotator cuff tear? *J Invest Med* 1996;**7**:364A.
Reason for exclusion: abstract only; later publication[33] included.

Liu SH, Henry MH, Nuccion S, Shapiro MS, Dorey F. Diagnosis of glenoid labral tears. A comparison between magnetic resonance imaging and clinical examinations. *Am J Sports Med* 1996;**24**:149–54.
Reason for exclusion: wrong population – labral tears.

Liu SH, Henry MH, Nuccion SL. A prospective evaluation of a new physical examination in predicting glenoid labral tears. *Am J Sports Med* 1996; **24**:721–5.
Reason for exclusion: wrong population – labral tears.

Loehr SP, Pope TL, Martin DF, Link KM, Monu JU, Hunter M, *et al*. Three-dimensional MRI of the glenoid labrum. *Skeletal Radiol* 1995;**24**:117–21.
Reason for exclusion: wrong population – labral tears.

Macfarlane GJ, Hunt IM, Silman AJ. Predictors of chronic shoulder pain: a population based prospective study. *J Rheumatol* 1998;**25**:1612–15.
Reason for exclusion: not diagnostic accuracy.

Mack LA, Matsen FA, Kilcoyne RF, Davies PK, Sickler ME. US evaluation of the rotator cuff. *Radiology* 1985;**157**:205–9.
Reason for exclusion: superseded by subsequent paper.[47]

Mamanee P, McFarland EG. The Neer impingement sign is associated with rotator cuff–glenoid rim contact. *Med Sci Sports Exerc* 1999;**31**:S237.
Reason for exclusion: insufficient data.

Manton GL, Schweitzer ME, Weishaupt D, Karasick D. Utility of MR arthrography in the diagnosis of adhesive capsulitis. *Skeletal Radiol* 2001;**30**:326–30.
Reason for exclusion: wrong population – adhesive capsulitis.

Marx RG, Bombardier C, Wright JG. What do we know about the reliability and validity of physical examination tests used to examine the upper extremity? *J Hand Surg* 1999;**24**:185–93.
Reason for exclusion: not assessing accuracy.

McCauley TR, Pope CF, Jokl P. Normal and abnormal glenoid labrum – assessment with multiplanar gradient-echo MR imaging. *Radiology* 1992;**183**:35–7.
Reason for exclusion: wrong population – labral tears.

Middleton WD, Reinus WR, Totty WG, Melson GL, Murphy WA. US of the biceps tendon apparatus. *Radiology* 1985;**157**:211–15.
Reason for exclusion: insufficient data.

Monu JU, Pruett S, Vanarthos WJ, Pope TL. Isolated subacromial bursal fluid on MRI of the shoulder in symptomatic patients: correlation with arthroscopic findings. *Skeletal Radiol* 1994;**23**:529–33.
Reason for exclusion: reference test – limited to index test positives.

Murrell GA,.Walton JR. Diagnosis of rotator cuff tears. *Lancet* 2001;**357**:769–70.
Reason for exclusion: insufficient data.

Neer CS. Impingement lesions. *Clin Orthop* 1983; **173**:70–7.
Reason for exclusion: not assessing accuracy.

Neumann CH, Petersen SA, Jahnke AH. MR imaging of the labral-capsular complex: normal variations. *Am J Roentgenol* 1991;**157**:1015–21.
Reason for exclusion: wrong population – labral tears.

Norwood LA, Barrack R, Jacobson KE. Clinical presentation of complete tears of the rotator cuff. *J Bone Joint Surg Ser A* 1989;**71**:499–505.
Reason for exclusion: study type – not accuracy.

Nuccion S, Hame SL, Chuan J, Seeger L. The accuracy of MRI and MRA in diagnosis of partial tears of the rotator cuff. *Med Sci Sports Exerc* 2001;**33**:S275.
Reason for exclusion: abstract only.

O'Brien SJ, Pagnani MJ, Fealy S, McGlynn SR, Wilson JB. The active compression test: a new and effective test for diagnosing labral tears and acromioclavicular joint abnormality. *Am J Sports Med* 1998;**26**:610–13.
Reason for exclusion: wrong population – labral tears.

Oh CH, Schweitzer ME, Spettell CM. Internal derangements of the shoulder: decision tree and cost-effectiveness analysis of conventional arthrography, conventional MRI, and MR arthrography. *Skeletal Radiol* 1999;**28**:670–8.
Reason for exclusion: not primary accuracy study (included in review of economic evaluations).

Peh WC, Campbell C, Kelly IG, Davidson JK. The role of computerized tomography in shoulder arthrography. *J R Coll Surg Edinb* 1994;**39**:190–5.
Reason for exclusion: ineligible index test – CTA.

Pisan M, Gerber C. Clinical examination of the rotator cuff. *Curr Orthop* 2000;**14**:155–60.
Reason for exclusion: study type – not accuracy.

Pope DP, Croft PR, Pritchard CM, Silman AJ. Prevalence of shoulder pain in the community: the influence of case definition. *Ann Rheum Dis* 1997;**56**:308–12.
Reason for exclusion: not accuracy study.

Rafii M, Firooznia H, Golimbu C, Minkoff J, Bonamo J. CT arthrography of capsular structures of the shoulder. *Am J Roentgenol* 1986;**146**:361–7.
Reason for exclusion: ineligible index test – CTA.

Railhac JJ, Sans NG, Galy-Fourcade D, Andrieu V, Dromer C, Fournie B. Adhesive capsulitis of the shoulder: a prospective study of the value of imaging techniques in 17 patients. *Radiology* 1998;**209P**:405.
Reason for exclusion: wrong population – adhesive capsulitis.

Roberts C, Seligson D. The accuracy of preoperative diagnostic ultrasonography of the shoulder. *Med Sci Sports Exerc* 1998;**5**:S260.
Reason for exclusion: abstract only.

Ryu KN, Lee SW, Rhee YG, Lim JH. Adhesive capsulitis of the shoulder joint: usefulness of dynamic sonography. *J Ultrasound Med* 1993;**12**:445–9.
Reason for exclusion: wrong population – adhesive capsulitis.

Seeger LL, Gold RH, Bassett LW, Ellman H. Shoulder impingement syndrome: MR findings in 53 shoulders. *Am J Roentgenol* 1988;**150**:343–7.
Reason for exclusion: reference test – limited to index test positives.

Seeger LL, Yao L, Gold RH. Diagnosis of glenoid labral tears: a comparison between magnetic resonance imaging and clinical examinations. *Am J Sports Med* 1997;**25**:141–4.
Reason for exclusion: wrong population – labral tears.

Sonnabend D, Hughes HJ, Giuffre B, Farrell R. Ultrasound assessment of shoulder pathology. In Surgery of the shoulder. 6th International Congress, Helsinki, Finland, 1995, pp. 13–17. Vastamaki, M and Jalovaara P, editors. International Congress Series (1085).
Reason for exclusion: not primary study.

Speer KP, Hannafin JA, Altchek DW, Warren RF. An evaluation of the shoulder relocation test. *Am J Sports Med* 1994;**22**:177–83.
Reason for exclusion: wrong population – instability.

Takagishi K, Itoman M, Araki T, Nishimura A, Yamamoto M, Ikeda T. MR imaging of recurrent dislocation and impingement syndrome of the shoulder. *J Orthop Surg* 1993;**1**:99–101.
Reason for exclusion: wrong population – labral tears.

Thomazeau H, Gleyze P, Lafosse L, Walch G, Kelberine F, Coudane H. Arthroscopic assessment of full-thickness rotator cuff tears. *Arthroscopy* 2000; **16**:367–72.
Reason for exclusion: wrong index test – arthroscopy.

Tirman PF, Bost FW, Steinbach LS, Mall JC, Peterfy CG, Sampson TG, *et al*. MR arthrographic depiction of tears of the rotator cuff: benefit of abduction and external rotation of the arm. *Radiology* 1994; **192**:851–6.
Reason for exclusion: wrong population – all throwing athletes.

Trattnig S, Breitenseher M, Rand T, Ba-ssalamah A, Schick S, Imhof H, *et al*. MR imaging-guided MR arthrography of the shoulder: clinical experience on a conventional closed high-field system. *Am J Roentgenol* 1999;**172**:1572–4.
Reason for exclusion: study type – not accuracy.

Tuite MJ, De Smet AA, Norris MA, Orwin JF. MR diagnosis of labral tears of the shoulder: value of T2*-weighted gradient-recalled echo images made in external rotation. *Am J Roentgenol* 1995;**164**:941–4.
Reason for exclusion: wrong population – labral tears.

Vecchio PL, King RH. Shoulder pain in a community-based rheumatology clinic. *Br J Rheumatol* 1995; **34**:440–2.
Reason for exclusion: not diagnostic accuracy study.

Wallny T, Wagner UA, Prange S, Schmitt O, Reich H. Evaluation of chronic tears of the rotator cuff by ultrasound. A new index. *J Bone Joint Surg Br* 1999;**81**: 675–8.
Reason for exclusion: study type – not accuracy.

Wilson AJ, Totty WG, Murphy WA, Hardy DC. Shoulder joint: arthrographic CT and long-term follow-up, with surgical correlation. *Radiology* 1989;**173**:329–33.
Reason for exclusion: ineligible index test – CTA

Yegudin-Ash J, Afzal M, Lee D. A comparative study of ultrasonography with magnetic resonance imaging in patients with painful shoulder using arthroscopy as a gold standard. *Arthritis Rheum* 1998;**9**:S260.
Reason for exclusion: insufficient data – abstract only.

Appendix 5

Clinical examination: detailed study methods

Study details	Eligibility criteria	Details of clinical examination	Reference test and test interpretation
Calis, 2000[30] Physical med. and rehab. clinic Turkey (years not reported) Prosp1 Consecutive Popl: 120 patients/ index: 125 shoulders/ ref. test: 125 shoulders *Mean age*: 51.6 (SD 13.9) *Sex*: 40% male *Outcomes*: IS[a] *Prevalence*: 70%	Patients aged 18–70 referred from rheumatology, orthopaedics or who 'applied' directly. Excluded if: inflammatory or systemic disease; acute traumatic conditions; postoperative conditions; or neck or elbow disorders [a] Also determined the accuracy of clin. exam. at distinguishing stages 1, 2 and 3 SIS compared with MRI (see Appendix 9 for MRI)	*Tests evaluated individually and in combination* Neer's test: scapular rotation is prevented and arm forced to elevation at an angle between flexion and abduction Hawkins test: arm is flexed to 90° then forced into internal rotation Horizontal adduction test: arm is forced to adduction towards the shoulder while the elbow is flexed Painful arc test: pain occurs between the angle of 60 and 90° of abduction Drop arm test: patient abducts shoulder to 90° then lets arm down slowly Yergason test: elbow flexed to 90° and the forearm pronated. With the wrist held, the patient has actively to supinate against resistance Speed test: elbow extended and forearm supinated, forward elevation of the humerus to 60° resisted	*Ref. test*: Subacromial injection. Details provided *Test interval*: Not reported *Further investigations*: Patients firstly had plain X-ray and MRI scanning (MRI reported in Birtane paper[88]) *Diagnostic criteria*: Reported *Test interpreters*: 2 physicians with 4–8 years of experience in shoulder management *Ref. test interpretation*: Not reported – states 'experienced hands'
Hertel, 1996[31] Orthopaedics Switzerland (1992–3) Prosp2 Consecutive Popl: 100 patients/ index: 100 shoulders/ ref. test: 100 shoulders *Median age*: 51 (16–79) *Sex*: 74% male *Outcomes*: Any tear *Prevalence*: 72%; 55%[b] [b] Results reported separately for those with supraspinatus and subscapularis defects	Patients with unilateral subacromial impingement syndrome stages 1–3 who underwent open surgery/arthroscopy Excluded if: any impairment of passive range of glenohumeral motion	*Tests evaluated individually (first 3 for supraspinatus, last 2 for subscapularis)* External rotation lag sign: for supraspinatus and infraspinatus. Ability to maintain external rotation in elevation with elbow supported (elbow flexed to 90°, shoulder at 20° elevation and near max. external rotation) Drop sign: detects infraspinatus. Ability to maintain external rotation with arm at 90° elevation, almost full external rotation and elbow flexed to 90° Internal rotation lag sign: detects subscapularis. Ability to maintain almost max. internal rotation, elbow flexed to 90°, shoulder in 20° elevation and 20° extension, dorsum of hand lifted from lumbar region until almost full internal rotation Jobe sign: Patient seated with back to the physician who holds the arm at 90° elevation and at almost full internal rotation with the elbow flexed at 90°. Patient asked to actively maintain position as the physician releases the wrist while supporting the elbow Lift-off signs: tests subscapularis. No details	*Ref test*: Open surgery/arthroscopy. No details provided. *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Reported. Patients divided into 2 overlapping groups on basis of operative diagnosis: those with supraspinatus and infraspinatus disease (87) and those with subscapularis disease (53). Results not reported for all 100 patients together *Test interpreters*: Not reported *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Details of clinical examination	Reference test and test interpretation
Itoi, 1999[32] Orthopaedic institute Japan (1996–7) Prosp1 (implied from text) Consecutive Popl: 136/index: 143 shoulders/ref. test: 143 shoulders *Mean age*: 43 (13 to 80) *Sex*: 77% male *Outcomes*: FT tear *Prevalence*: 24%	Patients with various shoulder symptoms Of 35 patients found to have FT RCTs: 19 supraspinatus only 11 supraspinatus and infrapinatus 5 supraspinatus, infrapinatus, and subscapularis	*Tests evaluated individually* Full can test: arm in 90° elevation in the scapular plane (scaption) and 45° external rotation (Kelly, 1986) Empty can test (Jobe test or supraspinatus test): arm in 90° elevation in the scapular plane and in full internal rotation (Jobe, 1982)	*Ref. test*: MRI. Details provided. *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Both tests positive when either (1) the examination induced pain, (2) the muscle strength was 4 or less (criteria provided) or (3) there was pain or muscle weakness or both *Test interpreters*: Not reported *Ref. test interpretation*: Not reported
Litaker, 2000[33] Orthopaedic practice USA (1990–4) Retrosp Consecutive Popl: 501 patients/ index: 448 shoulders/ ref. test: 448 shoulders *Mean age*: 57.4 (SD 12.6) *Sex*: 63% male *Outcomes*: Any tear *Prevalence*: 67%	Patients with suspected RCT who underwent arthrography Excluded if: indication for arthrography was adhesive capsulitis; recent fracture of affected joint; recent operative procedure of affected joint; radiographic evidence of osteoarthritis of glenohumeral joint Complete data available for 448/501. Those with incomplete data reported not to differ significantly in terms of demographic characteristics, historical or physical examination features	Tests evaluated individually plus expert diagnosis Passive elevation of shoulder: shoulder elevated to maximum distance with patient supine Passive external rotation: patient supine, elbow flexed to 90° and placed against side. Examiner rotates the humerus away from body Arc of pain test: passive elevation to 180°, arm actively descends in the abduction plane Impingement sign: patient supine, passive elevation to full, hand supinated and arm adducted against the ear then internally rotated Atrophy: both posterior shoulders are observed (supraspinatus and infraspinatus reported separately) Weakness: compared with contralateral arm Weakness with external rotation: arms extended at shoulder level and orientated 45° to the front with thumbs pointing down Weakness with external rotation: elbows flexed to 90° with thumbs up, shoulders rotated internally 20°. Examiner's hands outside patient's and patient directed to resist inward pressure Night pain Clinical expert's diagnosis: diagnostic impression of study physician. History of trauma	*Ref. test*: Double-contrast arthrography. No further details provided *Test interval*: Not described *Further investigations*: Not described *Diagnostic criteria*: Reported. Test results classified retrospectively on review of orthopaedist's dictated note *Test interpreters*: Single orthopaedic surgeon specialising in diagnosis and treatment of shoulder problems *Ref. test interpretation*: As above

continued

Study details	Eligibility criteria	Details of clinical examination	Reference test and test interpretation
Lyons, 1992[34] Specialist shoulder clinic UK (1989–91) Retrosp Not reported Popl: 45[c] patients/ index: 42 shoulders/ ref. test: 42 shoulders *Mean age*: Not reported *Sex*: 60% male *Outcomes*: Any tear *Prevalence*: 81%	Patients all referred to a specialist shoulder unit and operated on for suspected RCT [c] 2 case notes lost; 1 had no estimate of size of RCT	*Tests evaluated in combination only* Supraspinatus strength test: subjective estimate of force of resisted abduction at the wrist with elbow straight and arm in 20° abduction Infraspinatus strength test: estimating the resisted strength of external rotation at the wrist with elbow at 90° Rent test: RC palpated by the examiner from behind, with arm in internal and external rotation and then hyperextended	*Ref. test*: Surgery. No details provided *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Not described *Test interpreters*: Not reported *Ref. test interpretation*: Not reported
MacDonald, 2000[35] [d] Not reported Canada (years not reported) Prosp2 Consecutive Popl: 85 patients/ index: 85 shoulders/ ref. test: 85 shoulders *Mean age*: 40 (16–72) *Sex*: 73% male *Outcomes*: IS; bursitis; any tear *Prevalence*: 53; 28; 28%	Patients undergoing shoulder arthroscopy [d] Paper states data were prospectively entered on to a database, but were then compared retrospectively	*Tests evaluated alone and in combination*: Neer test: shoulder elevated to max. degree of internal rotation while stabilising scapula (Neer, 1983[111]) Hawkins test: forward passive flexion to 90° and max. internal rotation (Hawkins, 1980)[118]	*Ref. test*: Arthroscopy. No details provided *Test interval*: Not reported *Test interpreters*: Single operating surgeon *Ref. test interpretation*: Same operating surgeon *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Not reported. *Test interpreters*: Single operating surgeon *Ref. test interpretation*: Same operating surgeon

continued

Study details	Eligibility criteria	Details of clinical examination	Reference test and test interpretation
Read, 1998[36] Radiology/orthopaedics Australia (1993–4) Cannot tell design recruitment not reported Popl: 42 patients/ index: 42 shoulders/ ref. test: 42 shoulders *Mean age*: 44 (19–70) *Sex*: Not reported *Outcomes*: IS *Prevalence*: 81%	Patients undergoing shoulder surgery for suspected RC or long biceps tendon disease Either acute/chronic pain All had failed trial of conservative therapy or had acute injury No patients excluded on basis of age or technical difficulty	*Tests evaluated in combination only* Based on site of pain, limitation of functional activities and night ache. Physical examination included assessment of: positive impingement sign, greater tuberosity tenderness with or without subacromial crepitus and RC weakness. Patients also underwent ultrasound (US) (see Appendix 6). Although surgeon not blinded to US findings, surgery always undertaken on the basis of clinical data alone. NB: US findings would have helped in diagnosis and planning of treatment	*Ref. test*: Arthroscopy/surgery – open surgery if full tear suspected; arthroscopic for other cases. No further details provided *Test interval*: From US to surgery: mean 8.8 weeks, range 1 day to 11 months *Further investigations*: Not reported *Diagnostic criteria*: Not reported *Test interpreters*: Clin. exam. by single orthopaedic surgeon. US by single radiologist *Ref. test interpretation*: One surgeon
Wnorowski, 1997[37] Orthopaedic centre USA (1990–4) Retrosp Recruitment not reported Popl: ?[e] patients/ index: 39 shoulders/ ref. test: 39 shoulders Median age: 30 (20–75) *Sex*: Not reported Outcomes: Any tear Prevalence: 35%	Patients seen for a shoulder problem (and who underwent MRI) The majority of cases were relatively difficult, referrals from other surgeons or diagnostic dilemmas (the primary diagnosis was unclear after clin. exam.). [e] 20% of those undergoing arthroscopy during time period underwent MRI and were included Only those who had not already undergone MRI when referred (n = 31) were included in evaluation of clin. exam.	*Clinical examination*: No details Patients also underwent MRI (see Appendix 9)	*Ref. test*: Arthroscopy and subacromial bursal evaluation of the cuff. Where MRI indicated an FT tear but arthroscopy no or partial tear, an arthroscopic arthrogram was performed to confirm or negate an FT tear. It is not clear whether the result of the latter contributed to the establishment of the reference diagnosis *Test interval*: Mean interval between MRI and ref. test 2 months, median 12 weeks. Two patients had a delay of 2 years, one of whom had an FT tear *Further investigations*: Not reported *Diagnostic criteria*: Not reported *Test interpreters*: 12[f] radiologists in the CCS plus experienced musculoskeletal radiologist (ESS) and one author. Not clear whose clinical diagnosis was used – paper refers to 'primary clinical working diagnoses' but not clear if these were the diagnoses accompanying the patient on referral, or that of the ESS [f] 4 examined first 31 patients, then 1 each for last 8 patients *Ref. test interpretation*: One author

continued

Study details	Eligibility criteria	Details of clinical examination	Reference test and test interpretation
Wolf, 2001[38] Medical Center USA (1999–2000) Retrosp Consecutive Popl: 109 patients/ index: 109[f] shoulders/ ref. test: 109 shoulders *Mean age*: 51.2 (29–86) *Sex*: 61% male *Outcomes*: FT tear *Prevalence*: 42% [f] 71 underwent MRI. No details provided as to what criteria were used to decide whether a patient needed an MRI or not	Patients undergoing arthroscopy for diagnoses relating to shoulder pain and weakness	*Single test* Rent test: transdeltoid palpation performed anterior to the anterior margin of the acromiom through the deltoid. Forearm held with elbow flexed to allow rotational control in order to manoeuvre the arm while the examiner's other hand is used for palpation (Codman, 1911[119]; McLaughlin, 1962[120]) *MRI* No details given (reported in Appendix 9). Results of the MRI were evaluated independently of the clinical examination to avoid bias	*Ref. test*: Arthroscopy. No details provided *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Described. Retrospective chart review used to document clinical, surgical and MRI findings *Test interpreters*: Single surgeon *Ref. test interpretation*: As above
Zaslav, 2001[39] Advanced orthopaedic centres USA (years not reported) Prosp2 Consecutive Popl: 115[g] patients/ index: 110 shoulders/ ref. test: 110 shoulders *Mean age*: 44 (17–76) *Sex*: 59% male *Pain duration*: 10.9 months (range 2 months–4 years) *Outcomes*: Non-outlet IS *Prevalence*: 24% [g] 4 had negative impingement sign; 1 had avascular necrosis	Patients with a positive Neer overhead impingement sign who underwent arthroscopic shoulder surgery All patients had failed an average of 16 weeks conservative treatment (range 2–25 months). 2 subjects had both had previous open RC repairs	*Single test* Internal rotation resistance strength test (IRRST): arm positioned in 90° abduction in the coronal plane. Patient maximally resists external rotation then internal rotation with arm in 90° external rotation and 85° internal rotation. Test aims to distinguish those whose impingement is due to subacromial pathology (outlet impingement) from those in whom it is due to intra-articular pathology or microinstability with secondary impingement (non-outlet impingement) (new test)	*Ref. test*: Arthroscopic surgery. No details provided *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Reported. NB: those with non-outlet impingement considered test positive. Those with classical outlet impingement were test negative *Test interpreters*: Before arthroscopy, standard shoulder physical examination form was completed by one assistant, either by direct physical examination or review of the surgeons pre-op. evaluation *Ref. test interpretation*: Surgeon

FT, full thickness; IS, impingement syndrome; PT partial thickness.

Appendix 6

Ultrasound: detailed study methods

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Ahovuo, 1989[73] Radiology/orthopaedics Finland (years not reported) Cannot tell design Recruitment not reported Popl: NR patients/ index: 88 shoulders/ ref. test: 88 shoulders *Mean age*: 47.3 (20–84) *Sex*: Not reported *Outcomes*: FT *Prevalence*: 32%	Patients with chronic pain and dysfunction considered to be due to lesions of the tendons of the shoulder	*Scanner*: 7.5-MHz real-time linear-array transducer *Technique*: Static transverse and longitudinal scans of the tendons of the rotator cuff and biceps tendon obtained with the US technique as described by Mack *et al*., 1985[46] Both shoulders examined	*Ref. test*: Single-contrast arthrography. No further details *Test interval*: Not reported; US always performed before arthrography *Further investigations*: 15 patients underwent surgery because of shoulder pain and dysfunction. Results were combined with the arthrographic results to give reference standard result (all surgical results verified existing arthrographic findings) *Diagnostic criteria*: Criteria for diagnosis provided (Crass, 1984[80]; Middleton, 1985[121]; Ahovuo, 1986) *Test interpreters*: Not reported *Ref. test interpretation*: Not reported
Arslan, 1999[50] Radiology Turkey (years not reported) Retrosp Recruitment not reported Popl: 350 patients/ index: 102 (105 shoulders)/ ref. test: 102 shoulders *Mean age*: 46 (21–73) *Sex*: 53% male *Outcomes*: Any *Prevalence*: 49%	Patients with physical examination suggestive of RC injury Excluded if prior shoulder surgery	*Scanner*: 7.5-MHz linear-array transducer *Technique*: Longitudinal and axial scans performed. Standard techniques for US of the SA/SD bursa and biceps tendon were used (Middleton, 1986[41]; Farin, 1990[52]; Mack, 1985[46]; van Holsbeeck[78])	*Ref. test*: Double-contrast arthrography. No further details *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria reported (techniques used were referenced) *Test interpreters*: Two staff radiologists with special training in shoulder sonography. Sonographic findings reported using the department's Shoulder Sonogram Report Form. Sonographic reports and sonograms were retrospectively interpreted *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Brandt, 1989[70] Radiology/orthopaedics US (1987–88) Prosp3 Recruitment not reported Popl: 98 patients/ index: 98[a] shoulders/ ref. test: 62 shoulders arthrography; 38 surgery[b] [b] Various ref. tests used, only those undergoing arthrography included in analyses *Mean age*: [a]52 (range 20–73) *Sex*: 78% male (of 58 in final analysis) *Outcomes*: Any *Prevalence*: 48%	Patients with physical examination suggestive of RC injury Referred for arthrography and/or shoulder sonography Discussion states that population largely consists of 'problem patients'	*Scanner*: 7.5-MHz transducer (5 pts) or 5-MHz linear-array transducer *Technique*: Used technique described by Mack, 1985.[46] [b] 4 patients excluded from analyses owing to technically inadequate sonograms	*Ref. test*: Arthrography (62). 34 underwent arthrography alone, 28 had arthrography later followed by surgery and further 10 had surgery alone (results presented separately for subgroup undergoing surgery). No further details. Note: 26 had no ref. test (excluded from analyses) *Test interval*: 62 patients underwent both tests on same day (4 later excluded) *Further investigations*: Not reported *Diagnostic criteria*: Seven sonographic criteria as described by one or more previous authors (Crass, 1984[80]; Middleton, 1984[122], 1985[48], 1986[41]; Bretzke, 1985[49]; Mack, 1985[46], 1988[123]; Crass, 1986[124]) were recorded and analysed for the study. Authors then analysed each of 7 original criteria individually to determine their diagnostic value. They found the presence of central echogenic bands (Middleton, 1985[48], 1986[41]) and echogenic foci within the rotator cuff (Crass, 1986[124]; Bretzke, 1985[49]) to be highly subjective and of limited usefulness. Then compared accuracy for detection of RCT using all seven criteria to using only five of the criteria *Test interpreters*: Single experienced sonologist *Ref test interpretation*: Experienced orthopaedic radiologist
Brenneke, 1992[51] Not reported US (1987–90) Cannot tell design Consecutive Popl: 120 patients/ index: 120 shoulders/ ref. test: 120 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any; FT; PT *Prevalence*: 54; 32; 23%	Patients undergoing pre-operative US Discussion reports that all patients had significant shoulder dysfunction unrelieved by conservative means	*Scanner*: 5-MHz high resolution linear-array section scanner *Technique*: Technique used described by Mack, 1988.[123] Indeterminate results appear to have been excluded – states that those with positive or negative ultrasound included	*Ref. test*: Arthroscopy. No further details *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis of complete or partial RCT were provided (Mack, 1988[123]) *Test interpreters*: Radiologist *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Burk, 1989[71] Radiology/orthopaedics USA (years not reported) Prosp1 Recruitment not reported Popl: 41 patients/ index: 38 shoulders MRI/ ref. test: 23 shoulders US/38 3 excluded owing to claustrophobia *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 39%	Patients referred for evaluation of possible RCTs	Evaluated MRI and ultrasound (see Appendix 9 for MRI details) *Scanner*: 5-MHz real-time phased-array linear scanner *Technique*: Images obtained in longitudinal and transverse planes. Technique used by Mack, 1985[46]	*Ref. test*: Arthrography. Details provided *Test interval*: Sequentially and same day *Further investigations*: 16 also underwent surgery. Subgroup of surgical results presented separately. Not clear if surgical results also contributed to estimation of accuracy when arthrographic results used as ref. test *Diagnostic criteria*: Criteria reported. Assume based on Mack, 1985[46], but also reference Crass, 1988 to justify not including focal areas of hyperechogenicity as indicative of RCT[125] *Test interpreters*: Two sonographers with combined experience in shoulder sonography of 94 patients before beginning this study *Ref. test interpretation*: Arthrography by experienced musculoskeletal radiologist who was not involved in the interpretation of the MR images
Chiou, 1999[76] Radiology/orthopaedics Taiwan (years not reported) Cannot tell design Recruitment not reported Popl: 200 patients/ index: 200 shoulders/ ref. test: 55 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 73%	Patients with shoulder pain	*Scanner*: 7- or 10-MHz linear probe *Technique*: Not described This was a preliminary study reporting experiences with high resolution US in a number of applications including shoulders, which may explain limited details	*Ref. test*: Arthrography and/or surgery. No further details *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for full/partial tears given (Chiou, 1996[126]; van Holsbeeck, 1995[78]). *Test interpreters*: Not reported *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Crass, 1988[40] Radiology/orthopaedics USA (years not reported) Retrosp Other Popl: 500 patients/ index: 500 shoulders/ ref. test: 124[c] shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 54%	Patients referred for ultrasound of RC 100 were referred from practitioners in the community, but comprehensive follow-up available on only a few of this group	*Scanner*: 10-MHz high-resolution real-time equipment *Technique*: Positioned with arms in front of the chest and shoulders shrugged and also in a hyperextended internal rotation with the arm placed behind the back. Technique described in Crass, 1987[127] [c] Indeterminate US results excluded (16). Authors included only 108 patients in results, but reported arthrography results in 124 – have therefore estimated accuracy including 16 indeterminates as either test positive or test negative	*Ref. test*: Surgery (124). No further details. 165 also underwent arthrography but results for US vs arthrography are not reported *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis previously reported (Crass, 1984[80], 1985; Bretzke, 1985[49]) *Test interpreters*: 'Physician examiner' not described further *Ref. test interpretation*: Not reported
De Muynck, 1994[57] Physical medicine/orthopaedics Belgium (years not reported) Cannot tell design Recruitment not reported Popl: NR patients/ index: 50 shoulders/ ref. test: 50 shoulders *Mean age*: 52.9 (26–77) *Sex*: 66% male *Outcomes*: Any *Prevalence*: 71%	Patients with a clinically diagnosed presumptive RC rupture	*Scanner*: 7.5-MHz linear array transducer *Technique*: Followed a standardised protocol based on literature data (Mack, 1985[46]; Crass, 1987[127]; Bretzke, 1985[49])	*Ref. test*: Arthrography (38)/arthroscopy (5)/arthrotomy (9). No further details. Results presented separately per ref. test but cannot be combined as two patients seem to have undergone more than one ref. test (results add up to 52) *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria based on those described in the literature (Katthagen, 1988[128]; Crass, 1988[129]): focal thinning or sonographic absence of tendons. Based on changes in shape rather than irregularities in echogenicity as used by Middleton, 1986[130] and Mack, 1988[131] *Test interpreters*: Not reported *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Drakeford, 1990[68] Orthopaedics/radiology US (1986–7) Prosp2 Recruitment not reported Popl: NR patients/ index: 50 shoulders/ ref. test: 50 shoulders *Mean age*: 51 (24–69) *Sex*: 69% male *Pain duration*: mean 5 months (range 3–9 months) *Outcomes*: Any *Prevalence*: 24%	Patients with signs and symptoms consistent with a chronic impingement syndrome non-responsive to conservative treatment None had had previous shoulder surgery	*Scanner*: 7.5-MHz transducer Technique: Four transverse and 2 longitudinal static views followed by real-time scanning of the RC. Technique used not referenced	*Ref. test*: Arthrography. Standard double-contrast arthrography. No further details *Test interval*: Not reported *Further investigations*: Some patients had lidocaine injection test *Diagnostic criteria*: 'Standard' interpretations of RC sonograms were used (Bretzke, 1985[49]; Crass, 1986[124]; Gristina, 1987[132]; Mack, 1985[46]; Middelton, 1984[122], 1985[48], 1986[130]). Looking for echogenicity *Test interpreters*: A radiologist and an orthopaedic surgeon. Read real time and static views together. Static images also interpreted separately by 2 experienced radiologists (results not presented). Final sonogram results were based on two of the authors independently agreeing on 39/50 sonograms and consulting with each other to agree on the additional 11 *Ref. test interpretation*: Performed by the radiologist and interpreted by all four authors independently. Final arthrogram results were based on all 4 authors independently agreeing on 46/50 and three of the four agreeing on the remaining 4
Farin, 1990[52] Finland (1985–8) Cannot tell design Recruitment not reported Popl: 381 patients/ index: 381[d] shoulders/ ref. test: 102 shoulders *Mean age*: 45 (18–71) *Sex*: [d]58% male *Outcomes*: IS *Prevalence*: 36%	Patients with shoulder impingement syndrome	*Scanner*: 7.5-MHz real time linear array scanner *Technique*: Examination described – did not differ from that used in earlier studies (e.g. Crass, 1987[127]; Middleton, 1986[41]; Mack, 1985[46]; Crass, 1985[133]) and always included static and dynamic portions	*Ref test*: Surgery. No further details *Test interval*: Not stated *Further investigations*: Not stated *Diagnostic criteria*: Two specific criteria for the sonographic diagnosis of impingement syndrome were developed: fluid collection in the subacromial subdeltoid bursa and fluid in the bursal system that gradually distended the bursa and pooled laterally to the subdeltoid portion while the arm was elevated. Non-specific but suggestive findings were thickening or thinning of RC and hypo- or hyperechogenicity of the RC without bursal fluid collection *Test interpreters*: Single radiologist. Sonograms interpreted prospectively though this does not necessarily mean study was prospective in design *Ref. test interpretation*: Not stated

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Gratz, 1998[43] Germany (1996–97) Prosp2 Recruitment not reported Popl: NR patients/ index: 17 shoulders/ ref. test: 17 shoulders *Mean age*: 50 (40–65) *Sex*: 52% male *Outcomes*: Any *Prevalence*: 35%	Patients with clinically and sonographically suspected RC lesions 17 patients were examined 22 times, as 3 were suspected of having re-rupture of the same joint. Have only extracted results for 17 (using first reported reading for those who underwent more than one examination)	Main aim was to evaluate arthroscintigraphy but all patients also underwent US and arthrography *Scanner*: Not reported *Technique*: Not reported Indeterminate results reported and included in data extraction first as positive then as negative	*Ref. test*: Arthrography. Details provided *Test interval*: Not reported *Further investigations*: Patients underwent a variety of additional investigations: arthroscintigraphy (17), MRI (4), arthroscopy (8) and bone scans *Diagnostic criteria*: Criteria described (no reference given). For incomplete RC tear: local thinning of the rotator cuff, visualised as an area with only poor US echo. For complete RC tear: areas with no US echos at all. *Test interpreters*: Not stated *Ref. test interpretation*: Not stated
Hodler, 1988[79] Switzerland (1986–87) Retrosp Consecutive Popl: 180 patients/ index: 180 shoulders/ ref. test: 50[e] shoulders (51 shoulders) surgery, 39 arthrography *Mean age*: [e]19–75 *Sex*: [e]76% male *Outcomes*: Any *Prevalence*: 69% of 51 [e] Reported characteristics are for 50 patients.	Patients with clinical suspicion of RC tears Of the 130 who did not undergo ref test, some were stated to be awaiting surgery and the remainder underwent conservative therapy only due to 'minor clinical symptoms'	*Scanner*: 5-MHz (12 pts) or 7.5-MHz (38 patients) linear phased-array transducer *Technique*: Static and dynamic imaging were performed	*Ref. test*: Surgery in 51 cases, 39 of whom also underwent double-contrast arthrography. No details provided. Arthrographic results were integrated into report to referring physician *Test interval*: Patients usually received surgery within 2 months of ultrasonography *Further investigations*: Mention of arthroscopy or electromyography and tomography in addition to arthrography *Diagnostic criteria*: Criteria used for original (i.e. prospective) US interpretation not presented, but Bretzke, 1985[49] and Rapf, 1986[134] were referenced. Authors then went on to examine retrospectively the US studies and compare them with the surgical results to determine criteria indicative of FT and partial-thickness (PT) tears and these criteria are presented: FT tear if outer border of RC was concave instead of convex, if outer border was completely absent or if hyperechoic area in the RC; PT tear diagnosed if the outer border of the RC was straight instead of convex. We have extracted results for the original interpretation *Test interpreters*: Three radiologists *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Hodler, 1991[59] US (years not reported) Prosp2 Recruiment not reported Popl: 23 patients (24 shoulders)/index: 24 shoulders/ref. test: 24 shoulders *Mean age*: 57.8 (39–84) *Sex*: 43% male *Outcomes*: FT *Prevalence*: 63%	Patients with chronic shoulder pain and suspected RCT Pathogenesis other than degenerative disease (11 had rheumatoid arthritis, 3 chronic pain after trauma, 3 calcifying tendinitis and 6 degenerative changes)	Evaluated MRI and US (see Appendix 9 for MRI details) *Scanner*: 7.5-MHz phased-array transducer *Technique*: Supraspinatus and infraspinatus tendon routinely examined longitudinally and transversely. With the exception of the subscapularis, no dynamic study was performed	*Ref. test*: Arthrography. Details provided *Test interval*: Not reported *Further investigations*: Not stated *Diagnostic criteria*: Limited details of criteria provided (no reference but compare results with Mack, 1985[46] and Middleton, 1986[41]). Criteria for tear: complete loss of the cuff substance or a focal thinning *Test interpreters*: Single radiologist with special skill in sonography. Findings documented on hard copies in a standardised manner *Ref. test interpretation*: Not reported
Kurol, 1991[44] Radiology/orthopaedics Sweden (1986–90) Cannot tell design Recruitment not reported Popl: NR patients/ index: 58 shoulders/ ref. test: 58 shoulders *Mean age*: 46 (24–67) *Sex*: 46% male *Outcomes*: Any *Prevalence*: 41%	Patients with chronic subacromial pain selected for surgery	*Scanner*: 7.5-MHz linear array transducer *Technique*: Static and real-time exams conducted. Used technique of Middleton, 1984[122] and Crass, 1984[80]	*Ref. test*: Surgery. Partial anterior acromioplasty was performed according to Neer, 1972 *Test interval*: Not stated *Further investigations*: Conventional radiography and/or diagnostic subacromial injection *Diagnostic criteria*: Criteria described (no reference given). Ultrasonogram considered pathologic if (1) there was focal or generalised change in echogenicity, (2) thinning of the cuff, (3) no cuff substance. RCT diagnosed if 1, 2 or 1 and 2 present *Test interpreters*: Not stated. Thus, arthrography was not done, and the US findings were not considered by the surgeon prior to the decision to operate *Ref. test interpretation*: Not stated

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Mack, 1988[47] Radiology/orthopaedics US (1983–85) Cannot tell design Consecutive Popl: 134 patients (139 shoulders)/index: 139 shoulders/ref. test: 139 shoulders (only 99 reported here) *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 49%	Patients with histories and physical examinations suggestive of RCT Excluded those with prior surgery	*Scanner*: 7.5-MHz mechanical sector scanner (first 43 examinations); 5- or 7.5-MHz linear array scanners (remainder) *Technique*: As described by Mack, 1985.[46] Patient seated on low rotating stool. Both shoulders imaged routinely. Images made sequentially in seven reference positions and recorded on multiformat film. The dynamic portion of the examination is recorded on videotape	*Ref. test*: Arthrography. 99 underwent double contrast arthrography, 90 had surgical exploration of cuff and 50 had both arthrography and surgical cuff exploration *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: RCT diagnosed as the absence of the normal tendons on static and dynamic images. Secondary findings often associated with RCT included joint effusion, narrowing of the subacromial space, irregularity of the humeral joint surface and concavity of the superior margin of the supraspinatus tendon. Changes in tendon echogenicity and the relative brightness of the tendons were recorded but not considered in the interpretation. These changes in echo level unreliable indicator of cuff tears *Test interpreters*: One of two staff radiologists experienced in US. Blinded to results of arthrography *Ref. test interpretation*: Experienced orthopaedic radiologist
Martin-Hervas, 2001[66] Orthopaedic unit (Univ. hospital) Spain (1998) Prosp3 Consecutive Popl: 140 patients (shoulders)/index: 140 US shoulders, 61[f] MRI/ref. test: 72[g] shoulders *Mean age*: Not reported *Sex*: [f]41% male [g] 11 unable to have MRI as claustrophobia, metal implants or pacemaker *Outcomes*: Any; FT; PT *Prevalence*: 56; 43; 13%	Patients with shoulder pain and limited movement fulfilling criteria for orthopaedic surgery protocol (n = 72). Only 61 undergoing both US and MRI included in analyses	Evaluated MRI and ultrasound (see Appendix 9 for MRI details) *Scanner*: 7.5-MHz high-resolution linear electronic transducer *Technique*: Transverse, longitudinal and posterior planes. Focus on suprapinatus tendon	*Ref. test*: Arthroscopy (40)/surgery (10)/both (11). No details provided *Test interval*: <6 months in all cases *Further investigations*: Not reported *Diagnostic criteria*: Individual criteria reported, but not clear how these contributed to diagnosis. Frequency of individual signs appear to have been classified retrospectively. Signs diagnostic of FT tear: complete absence of the tendon (4.9%), focal atrophy (32.8%), concave border (24.6%), liquid-filled hypoechoic bands (1.6%) and/or lineal hyperechoic bands (3.3%) (Taboury, 1995; Thain, 1999; van Holsbeeck, 1991). Signs 'suggestive' of tear included: heterogeneous tendon with hypoechoic areas (>3 mm) that do not reach both sides of the tear (sign of an intratendinous tear) (4.9%) and an irregular or indented border (a sign of a partial tear) (19.7%) Other findings visible in partial and full tears: effusions in the peribicipital region (19.7%), the subacromial bursa (29.5%) or the joint (9.8%), heterogeneous tendon (13.1%) or calcifications (27.9%) *Test interpreters*: Two musculoskeletal radiologists *Ref. test interpretation*: Not described

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Middleton, 1986[41,48] Radiology US (1983–4) Prosp3 Recruitment not reported Popl: 106 patients/ index: 106 shoulders/ ref. test: 106[h] shoulders *Mean age*: 47 (12–84) *Sex*: 71% male *Outcomes*: Any *Prevalence*: 34%	Patients referred for arthrography	*Scanner*: 10-MHz high-resolution scanner *Technique*: Standardised to include transverse scans of the bicipital groove, the distal area of the biceps tendon and the intra-articular part of the biceps tendon followed by longitudinal scans of the long axis of the tendon. Using the biceps tendon and groove as a landmark, the supraspinatus tendon was localised and scanned in both the longitudinal and transverse planes [h] Indeterminate results excluded from final analysis ($n = 6$)	*Ref. test*: Arthrography (standard double-contrast technique). Details provided. Authors admit grading is relatively subjective *Test interval*: US conducted immediately before arthrography *Further investigations*: None reported *Diagnostic criteria*: Scans scored as positive, negative or indeterminate. Once results of arthrography were known, they seem to have identified sonographic appearances suggestive of tear: focal thinning (11 patients all with RCT); complete non-visualisation of cuff (10 all with RCT); focal discontinuity in homogeneous echogenicity of cuff without thinning (10, 5 with RCT); central echogenic band within region of RC (6, 5 with RCT) *Test interpreters*: Two radiologists *Ref. test interpretation*: Arthrogram report retrospectively reviewed by two authors to standardise comparison of with US
Miller, 1989[72] Radiology USA (1986–7) Prosp1 Recruitment not reported Popl: 57[i] patients/ index: 56 shoulders/ ref. test: 56 shoulders *Mean age*: 55 (24–79) *Sex*: 54% male *Outcomes*: Any *Prevalence*: 46%	Patients referred for ultrasound of the shoulders because of clinical suspicion of rotator cuff pathology	*Scanner*: 5-MHz real-time phased-array linear scanner *Technique*: used technique of Mack *et al.*, 1985[46] [i] 1 indeterminate result excluded. Discussion further states that in many cases ultrasound findings fell into a 'grey zone'	*Ref. test*: Double-contrast arthrography. No further details *Test interval*: Arthrography performed within 1 h of sonography *Further investigations*: 8 RCs were surgically repaired – surgical reports available on 4 *Diagnostic criteria*: Described (not clearly referenced, but appear to have used those of Mack, 1985)[46] The presence of any one of three major criteria determined the presence of an RCT: (1) total absence of the cuff, (2) sever distortion in the sonographic architecture of the cuff, (3) marked thinning of the cuff *Test interpreters*: Two sonologists (qualified with MDs) working in hospital radiology departments *Ref. test interpretation*: Same as for index test

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Misamore, 1991[58] Sports medicine centre USA (1988–9) Prosp3 Consecutive Popl: 138 patients/ index: 82 shoulders/ ref. test: 32[j] shoulders *Mean age*: All 35–65 *Sex*: 81% male *Outcomes*: Any; *Prevalence*: 84%	Patients over 35 years old with symptoms and physical findings consistent with chronic tendinitis or degeneration or a tear of the RC, including patients who had signs indicating stage II or stage III impingement Pain duration at least 1 year Acute injury, symptoms less than 1 year, previous shoulder surgery or associated shoulder disorders excluded	*Scanner*: Real-time US in two planes – no details of field strength *Technique*: Not stated [j] Only those with severe symptoms proceeded on to surgery ($n = 32$), another 50 had arthrography but were not severe enough for surgery, 26 had symptoms not severe enough to warrant arthrography or US and 30 more were excluded owing to age <35 (excluded as many younger patients have a non-degenerative type of tendinitis)	*Ref. test*: Arthroscopy/open surgery. Either diagnostic arthroscopy ($n = 2$), operative arthroscopy ($n = 5$) or open operation (open acromioplasty in 4, repair RC in 21) *Test interval*: Not reported *Further investigations*: Double-contrast arthrography with radiographs before and after exercise also evaluated against surgery. Clinical diagnosis of tendinitis supported by relief of symptoms on impingement tests (subacromial injection with local anaesthetic) *Diagnostic criteria*: Criteria reported (no reference given). Considered to be positive when an obvious defect localised to the tendon of the RC was seen or, alternatively, when there was no echo of the RC. An abnormality of echogenicity alone was not considered a sufficient criterion for a diagnosis of a tear *Test interpreters*: Two radiologists skilled in both US and arthrography. Both tests performed and interpreted by different radiologists neither of whom knew the findings of the other test. Also a retrospective review of the studies by two authors and the radiologists but no changes made in initial diagnosis *Ref. test interpretation*: One author
Naredo, 1999[60] Spain (years not reported) Prosp1 Consecutive Popl: 34 patients/index: 34 shoulders (36 shoulders)/ ref. test: 36 shoulders *Mean age*: 61 (36–75) *Sex*: 3% male *Pain duration*: Mean 7 months (14 days to 36 months) *Outcomes*: FT; PT; Ten *Prevalence*: 50; 36; 39%	Patients referred to rheumatology clinic with first flare of shoulder pain Excluded: previous trauma, chronic inflammatory arthritis	*Scanner*: 7.5-MHz linear phased-array transducer *Technique*: Real-time transverse and longitudinal scans taken. Details given	*Ref. test*: MRI, details provided *Test interval*: <2 weeks *Further investigations*: Minority of patients underwent surgery *Diagnostic criteria*: 14 categories used, based on those used in literature (van Holsbeeck, 1991,[138] 1992,[139] 1995[78]; Middleton, 1986[41]; Brtezke, 1985[49]; Mack, 1985[46]). Criteria for each was detailed. Accuracy reported for 10 separate outcomes. Only suprapinatus tears reported here. FT tear: non-visualisation of tendon or fibre discontinuity from humeral head to subacromial–sub-deltoid bursa or superior convexity instead of concavity PT tear: hyperechoic fibre discontinuity involving bursal or articular surface or intrasubstance hyperechoic defect or focal tendon thinning RC tendinitis: tendon hypoechogenicity or tendon thickening with or without internal hyper- or hypoechoic foci *Test interpreters*: Single rheumatologist experienced in US *Ref. test interpretation*: One musculoskeletal radiologist

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Nelson, 1991[64] Radiology/orthopaedics US (years not reported) Prosp1 Recruitment not reported Popl: 21 patients/index: MRI 21, ultrasound 19 shoulders/ ref. test: 21 shoulders *Mean age*: 42 (18–70) *Sex*: 76% male *Outcomes*: FT; PT *Prevalence*: 37%; 57%[k] [k] prevalences across all 21 patients	Patients with shoulder pain resistant to non-operative treatment History of shoulder pain >3 months before first consultation	Evaluated MRI and US (see Appendix 9 for MRI details) *Scanner*: 5-MHz linear-array transducer *Technique*: Imaging sequence obtained in longitudinal and transverse planes. Real-time interpretation Also evaluated CTA	*Ref. test*: Arthroscopy (18)/open surgery (1)/both (2). No details provided *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis stated (no reference given) but not clear how they differentiated between full and partial. Were looking for presence of any thinning, irregularity, calcification, hyperechoic or hypoechoic areas and collections of fluid [k] Indeterminate results not mentioned but may have been excluded as US reported for only 19/21 patients *Test interpreters*: Two radiologists 'with the necessary expertise'. Complete set of studies (MRI, US, CTA) for each patient was read by one radiologist whose report influenced choice of treatment (ref. test). All of the studies were also read by a second radiologist who was unaware of the results of the other studies, type of clinical, management and operative findings. US images were interpreted at the time of the study by a third radiologist. However, findings of the first radiologist were used as the final reading for the purpose of the study *Ref. test interpretation*: Not reported
Olive, 1992[45] Radiology USA (1987–8) Prosp3 Consecutive Popl: 89 patients/ index: 72 shoulders/ ref. test: 72 shoulders *Median age*: 50 (18–86) *Sex*: 56% male *Outcomes*: FT *Prevalence*: 40%	Patients with shoulder pain due to suspected RCTs	*Scanner*: 7.5-MHz, high-resolution scanner *Technique*: Standard protocol developed based on literature (Bretzke, 1985[49]; Crass, 1984[80], 1988[125]; Collins, 1987[135]; Mack, 1985,[46] 1988[47]; Middleton, 1984,[122] 1985,[48] 1986[41])	*Ref. test*: Arthrography. No details provided *Test interval*: In 64 patients sonogram immediately before arthrography. In remaining 8 arthrography performed first owing to scheduling problems; US undertaken at least 4 days later, no further details *Further investigations*: Not reported *Diagnostic criteria*: Used criteria established by Middleton, 1985.[48] Full tear: discontinuity in echogenicity of cuff, central echogenic band replacing normal homogeneous echogenicity, non-visualisation of RC. Incomplete tears or other shoulder abnormalities not related to the RC were considered negative *Test interpreters*: Designated staff radiologists; one of two interpreted each *Ref. test interpretation*: Staff radiologists. Number not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Paavolainen, 1994[61] Orthopaedics/radiology Finland (1986–90) Retrosp Recruitment not reported Popl: 428 patients/ index: 49 shoulders/ ref. test: 49 shoulders *Mean age*: 38 (24–76) *Sex*: 69% male *Other*: 29% had pain onset related to trauma *Outcomes*: FT *Prevalence*: 55%	Patients undergoing US who had shoulder pain or dysfunction secondary to degenerative changes and a PT or FT RCT. All failed conservative therapy Excluded: pain <6 months; non-shoulder source of pain; arthritis or instability of shoulder; previous shoulder surgery	*Scanner*: 7.5-MHz linear-array scanner *Technique*: Scans according to Mack, 1985.[46] Technique described Note: May be some overlap with Ahovuo, 1989.[73] Both studies conducted in same department (but years not reported for Ahovuo): 88 patients underwent US and arthrography, 15 of whom also underwent surgery – the latter could have been included in Paavolainen report	*Ref. test*: Open surgery: anterior acromioplasty, excision of any osseous prominence and of subacromial bursa. RC and biceps tendon examined *Test interval*: Not reported *Further investigations*: All patients also underwent single-contrast arthrography. *Diagnostic criteria*: Criteria reported (Mack, 1985[46]; Middleton, 1985[48]). RCT: local thinning or absence of echo of RC. Abnormal echogenicity alone (hyperechogenicity or hypoechogenicity) was not considered a sufficient criterion for RCT (Ahovuo, 1989[73]). Lesions of head of biceps also recorded *Test interpreters*: Two radiologists. Reported not blinded: US and arthrograms were initially conducted and interpreted independently (not clear if clinical info. available) and were later jointly reviewed subsequent to the operative procedure. Discussion states that no changes in the initial interpretations of the tests occurred once the operative findings were known *Ref. test interpretation*: No details
Pattee, 1988[69] Orthopaedics USA (1985–7) Cannot tell design Recruitment not reported Popl: NR patients/ index: 52 shoulders/ ref. test: 52 shoulders *Mean age*: 47 (28–71) *Sex*: 81% male *Pain duration*: Mean 24 months (3 months – 9 years) *Outcomes*: Any *Prevalence*: 67%	Patients with chronic shoulder pain, signs of an impingement lesion and clinical suspicion of a RCT, who failed to respond to conservative treatment for at least 3 months	*Scanner*: 7.5-MHz transducer *Technique*: Longitudinal and transverse images obtained	*Ref. test*: Arthroscopy: views of glenohumeral joint, posterior and superior surface of RC *Test interval*: Not stated *Further investigations*: Single-contrast arthrography performed in 32 patients. Note: only 11/22 tears noted arthroscopically were also detected arthrographically (improves if PT tears are excluded) *Diagnostic criteria*: Criteria provided (Middleton, 1986[41]; Bretzke, 1985[49]; Crass, 1984[80] in discussion). Classified as normal or tear. Tear present if areas of increased echogenicity, focal thinning, area of discontinuity, or complete non-visualisation of RC *Test interpreters*: Single radiologist *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Read, 1998[36] Radiology/orthopaedics Australia (1993–4) Cannot tell design Consecutive Popl: 42 patients/index: 42 shoulders/ref. test: 42 shoulders *Mean age*: 44 (19–70) *Sex*: Not reported *Outcomes*: FT; PT; IS *Prevalence*: 24; 31; 81%	Patients with acute/chronic shoulder pain who all had failed trial of conservative therapy or had acute injury and underwent shoulder surgery for suspected RC or long biceps tendon disease	Evaluated both ultrasound and clinical examination (see Appendix 5 for clin. exam. details) *Scanner*: 7.5-MHz linear array Technique: Static images routinely obtained according to methods described by Mack, 1985,[49] 1988[47] and Middleton, 1985,[48] 1986,[41] 1992.[136] Dynamic component used to identify any signs of impingement in the lateral acromion, anterior acromion and the coracoacromial ligament (similarly to Farin, 1995[137]) No patient excluded on basis of age or technical difficulty	*Ref. test*: Arthroscopy/open surgery (if full tear suspected). No details provided *Test interval*: US to surgery: mean 8.8 weeks; range 1 day to 11 months *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis provided (criteria presumably taken from Mack and Middleton). NB: RC tendinitis also diagnosed by ultrasound (in 15 cases) but accuracy results not presented, partly because surgical correlation difficult but also usually a significant delay between sonography and operation in these patients (mean 14.7 weeks). *Test interpreters*: Single radiologist *Ref. test interpretation*: One surgeon. Surgeon not blinded to US findings, surgery always undertaken on the basis of clinical data alone. NB: US findings would have helped in diagnosis and planning of treatment
Roberts, 2001[77] Orthopaedic surgery USA (years not reported) Prosp3 Recruitment not reported Popl: NR shoulders/index: 24 shoulders/ ref. test: 24 shoulders *Mean age*: 49 (37–75) *Sex*: 67% male *Outcomes*: Any; FT; PT *Prevalence*: 71; 42; 29%	Patients scheduled for shoulder arthroscopy	*Scanner*: 7.5-MHz transducer Technique: Full details provided. Scanned in transverse and longitudinal planes	*Ref. test*: Arthroscopy: bursal and articular surfaces examined *Test interval*: Not reported. States US performed 'prior to arthroscopy' *Further investigations*: 11 patients had previous diagnostic investigations (7 MRI and 6 arthrograms). Additional procedures performed at time of arthroscopy included subacromial decompression ($n = 20$) and mini-open RC repair ($n = 7$) *Diagnostic criteria*: Standard criteria used (Middleton, 1985[48]). FT tear: discontinuity in normal homogenicity of RC or non-visualisation of RC. PT tear: flattening of bursal surface of RC or hypoecholic area on longitudinal and transverse planes of articular surface of RC *Test interpreters*: Single orthopaedic surgeon with prior experience and training in diagnostic musculoskeletal ultrasonography. US interpreted prospectively and diagnosis recorded. Not masked to results of other diagnostic investigations or clinical findings; may have been more inclined to look for PT in those in whom they had already been diagnosed by other methods *Ref. test interpretation*: Not reported. Arthroscopist also aware of ultrasound findings again leading to potential for looking longer for a tear in those with positive US or overcalling areas of mild tendon fraying

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Shipley, 1985[65] Ultrasound department Canada (years not reported) Cannot tell design Recruitment not reported Popl: NR shoulders/ index: 12 shoulders/ ref. test: 12 shoulders *Mean age*: 65 *Sex*: Not reported *Outcomes*: Any *Prevalence*: 83%	Patients with suspected RCT referred by orthopaedic surgeon for arthrography	*Scanner*: With 5-MHz linear array (initially used 10-MHz water path transducer but not clear if any results were obtained using this) *Technique*: Scans performed in transverse and longitudinal planes	*Ref. test*: Arthrography: no details *Test interval*: US performed prior to arthrography *Further investigations*: Not reported *Diagnostic criteria*: Criteria reported (Middleton, 1985[48]). Criteria for detection of RCT: total absence of visualisation of the cuff; an area of increased echogenicity within the cuff; presence of echogenic bands running longitudinally through the cuff on the coronal scan *Test interpreters*: Single interpreter, qualifications not reported. US abnormality: determined retrospectively alongside arthrography. Authors trying to determine the appropriate diagnostic criteria to use rather than simply assessing the accuracy of the test *Ref. test interpretation*: No details
Soble, 1989[75] Radiology dept USA (1986–8) Cannot tell design Recruitment not reported Popl: 86? shoulders/ index: 75 shoulders/ ref. test: 75 shoulders *Mean age*: (17–85) *Sex*: Not reported *Outcomes*: Any *Prevalence*: 35%	Patients with suspected RCT referred by community orthopaedists	*Scanner*: 7.5-MHz scanner or a 5-MHz phased linear array *Technique*: Used technique as described in literature (Middleton, 1984,[122] 1986[41]; Mack, 1985,[46] 1988[47]; Crass, 1985,[133] 1986,[124] 1988[125]; Bretzke, 1985[49]). Scanned in transverse and longitudinal planes. Routinely image only supraspinatus and infraspinatus tendons	*Ref. test*: Arthrography: single contrast ($n = 46$) or double contrast ($n = 29$) at discretion of surgeon. Technique used as previously described in literature (ref. provided) *Test interval*: US performed immediately before arthrography *Further investigations*: Surgery also undertaken in 30 patients for RCT or other soft tissue abnormality. Found 2 additional tears during surgery that were missed by arthrography. One of these was detected at US and so should have been a TP rather than an FN. Have extracted arthrography results only *Diagnostic criteria*: Criteria given. Abnormal RC classified as consistent with tear (if fulfilled major diagnostic criteria) or cannot exclude tear (if any other abnormality found). Discussion states that the major diagnostic criteria used conform to those of Crass,1988[125] but are less restrictive than those of Mack, 1985,[46] 1988[47] *Test interpreters*: Radiologist. Familiarity with normal sonographic anatomy was obtained by scanning asymptomatic people in the department *Ref. test interpretation*: Surgeon

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Sonnabend, 1997[53] Australia (1991–4) Cannot tell design Consecutive Popl: 117 patients/ index: 117 shoulders/ ref. test: 117[l] shoulders *Mean age*: 49 (14–79) *Sex*: 71% male *Outcomes*: Any; FT; PT *Prevalence*: 63; 41; 22%	Patients with clinical diagnosis of impingement but no cuff tear who had failed 3 months of conservative management and underwent bilateral US followed by surgery [l] states that 117 patients were included but can only find results for 110 (in both papers)	*Scanner*: 7.5-MHz linear array transducer *Technique*: Longitudinal and transverse scans using both static and dynamic techniques. Images interpreted in real time	*Ref test*: Arthroscopy/surgery. No details provided. *Test interval*: Median time 3 months. Note that one of the missed tears was very large and was confirmed at surgery 5 months after US – suggest the lag time between ultrasound and surgery may be relevant *Further investigations*: Not reported *Diagnostic criteria*: Criteria used stated to be comparable to those previously reported (Weiner, 1993[56]; Vick, 1990[74]; Hodler, 1988[79]) *Test interpreters*: 2 experienced radiologists *Ref. test interpretation*: No details
Swen, 1998[62] Orthopaedics, rheumatology and radiology The Netherlands (1993–5) Prosp2 Consecutive Popl: 48 patients/ index: 48 shoulders/ ref. test: 48 shoulders *Mean age*: 55 (30–76) *Sex*: 58% male *Pain duration*: Mean 2 years (0.3–10 years) *Outcomes*: FT *Prevalence*: 46%	Patients awaiting surgery for based on clinical suspicion of RCT (based on marked difficulty in initiating abduction of the arm, with weakness and limitation of movement) All had chronic (⩾3 months) unilateral shoulder complaints without underlying inflammatory disease	*Scanner*: 7.5-MHz linear-array or 5.0-MHz curved array type *Technique*: US performed while patients were seated. Used technique of van Holsbeeck,1991[138] Also evaluated arthrography	*Ref. test*: Surgery: arthroscopic decompression (11) or open decompression (37) where only the bursal aspect of the cuff is inspected *Test interval*: Not clear *Further investigations*: All underwent double-contrast arthrograms within 6 weeks before surgery (study also compared arthrography and arthroscopy results) *Diagnostic criteria*: Used criteria as established by Wiener, 1993[56] and van Holsbeek, 1991.[138] FT RCT defined as a discontinuity in the RC, extending from the bursal to the humeral side of the rotator cuff *Test interpreters*: Single rheumatologist with experience in the technique *Ref. test interpretation*: 3 surgeons

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Swen, 1999[63] Rheumatology The Netherlands (years not reported) Prosp2 Consecutive Popl: 21 patients/ index: 21 shoulders/ ref. test: 21 shoulders *Mean age*: 54 (SD 12) *Sex*: 57% male *Pain duration*: Mean 2.3 years (0.3–8 years) *Outcomes*: FT *Prevalence*: 62%	Patients awaiting surgery for clinically suspected RCT based on marked difficulty in abducting the arm, weakness and limitation of movement, even following lidocaine injection All had unilateral non-inflammatory symptoms. In 4 patients shoulder complaints could be attributed to trauma Neurological origins of shoulder weakness were excluded	Evaluated MRI and ultrasound (see Appendix 9 for MRI details) *Scanner*: 7.5-MHz linear array and 5.0-MHz curved array (latter used for an overview of infraspinatus tendon and joint capsule) *Technique*: Method of van Holsbeeck (1991) was used with imaging in transverse and longitudinal planes. Sonography performed twice in 3 weeks May be some overlap of patients with Swen *et al.*, 1998[62]	*Ref. test*: Arthroscopy. Subacromial decompression performed in all cases. FT RCTs repaired through a deltoid-splitting mini-incision *Test interval*: Within 3 weeks *Further investigations*: None reported *Diagnostic criteria*: Van Holsbeeck's criteria (1991[138]) for FT RCT were used: discontinuity in the RC extending from the bursal to the humeral side of the RC. All RC tendons were evaluated, but only changes in the supraspinatus tendon were analysed in this study since RCTs almost always involve this tendon because of the 'critical zone' *Test interpreters*: Experienced radiologist and rheumatologist both performed US on all patients. Mean values of the 2 observers were used to estimate sensitivity/specificity reported in study (rounding means 2×2 data are not quite right), but results per reader also presented and extracted *Ref. test interpretation*: 1 experienced shoulder surgeon
Takagishi, 1996[54] Orthopaedics/radiology Japan (1987–94) Cannot tell design Recruitment not reported Popl: over 200 patients/ index: 122 shoulders/ ref. test: 122 shoulders *Mean age*: 51 (26–81) *Sex*: 63% male *Outcomes*: Any; FT; PT *Prevalence*: 48; 31; 16%	Patients with signs and symptoms suggesting RCT who ultimately required surgery	*Scanner*: 7.5-MHz linear-array transducer *Technique*: Real-time imaging. Used technique described by Crass, 1987[127] and Middleton, 1985[121]	*Ref. test*: Surgery: open anterior acromioplasty and repair of RC for almost all patients with RCT. Where an arthroscopic procedure was performed, RC observed through a small open incision *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for full or partial tears provided (not referenced but assume similar to those of Crass[127] and Middleton[121]). FT RCT: discontinuity and thinning of RC. PT RCT: dominant echogenic foci, hypoechogenic focus with an ecogenic focus and irregularity of outer surface of RC *Test interpreters*: One sonographer and one orthopaedic surgeon. Not clear whether all images read by both or whether each image read by one *Ref. test interpretation*: No details

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Teefey, 2000[55] Radiology and orthopaedics USA (1996–7) Retrosp Consecutive Popl: 98 patients (100 shoulders)/index: 100 shoulders/ref. test: 100 shoulders *Mean age*: 56 (14–82) *Sex*: 45% male *Pain duration*: Generally >3 months *Outcomes*: Any; FT; PT *Prevalence*: 80; 65; 15%	Patients with shoulder pain who had undergone pre-operative US followed by arthroscopy General indications for surgery and arthroscopy were: pain >6 months; no response to conservative therapy; for FT RCT indications were severe pain >3 months or recent loss of shoulder elevation or recent injury Authors note that patients with normal US and resolution of symptoms did not undergo arthroscopy and were not included in the study	*Scanner*: 7.5–10-MHz variable high-frequency linear-array, real-time image *Technique*: Followed a standardised protocol	*Ref. test*: Arthroscopy: findings recorded in a standardised manner. No details provided *Test interval*: Mean 60 days; range 1–417 days *Further investigations*: Where PT tear present or arthroscopic findings discrepant from those at US, arthroscopic bursal imaging was undertaken. Where an FT tear seen on US but not visualised at arthroscopy, an extensive PT tear was present and a mini-open deltoid split was performed to allow direct visualisation of the involved area of the RC and to verify arthroscopic findings. All FT tears repaired during this process *Diagnostic criteria*: Criteria for diagnosis provided (no reference given). FT RCT: non-visualisation of RC or when there was a focal defect in the RC created by a separation of the torn tendon ends when deltoid compressed against RC. PT RCT: minimal flattening of bursal side of RC or distinct hyperechoic or mixed hyper- and hypoechoic defect in both longitudinal and transverse plane at deep articular side of RC *Test interpreters*: Two radiologists with extensive experience in ultrasound. Although a retrospective study, the statistical analysis was based on the original interpretation of the US rather than on retrospective review of images *Ref. test interpretation*: Single orthopaedic surgeon. Arthroscopy unblinded but this was 'advantageous' to the patient as it allowed more focused evaluation of RC

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
van Holsbeeck, 1995[78] Orthopaedics/radiology USA [years not reported (2 year period)]] Cannot tell design Consecutive Popl: 67 shoulders, 52 patients/ index: 67 shoulders/ ref. test: 67 shoulders *Mean age*: 52 (23–86) *Sex*: 58% male *Outcomes*: Any; FT; PT; *Prevalence*: 73; 51; 22%	Patients with shoulder pain >3 months and findings of impingement on clinical examination	*Scanner*: 7.5-MHz linear array *Technique*: Both longitudinal and transverse images were obtained	*Ref. test*: Arthroscopy: glenohumeral and bursal. Patients considered for arthroscopy on the basis of clinical examination only (i.e. US result did not influence decision to undergo arthroscopy) *Test interval*: Mean 6 months; range 2 weeks – 11 months *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis of FT or PT were provided (no reference). PT RCT: two orthogonal imaging planes showed mixed hyper- and hypoechoic region or hypoechoic discontinuity in RC tendons. Lesions had to involve tendon on bursal or articular surface. FT RCT: tears extending through both articular and bursal surfaces. Focal thinning and loss of parallelism also indicative of FT RCT *Test interpreters*: One author (from Radiology Dept) *Ref. test interpretation*: 3 physicians. Arthroscopic and ultrasound results correlated using a videotape of the arthroscopic and hard copy US images. No blinding, may have meant that physicians searched harder for partial tears during arthroscopy in patients with positive ultrasound findings and not as hard in those with negative ultrasound findings
Van Moppes, 1995[67] Radiology/orthopaedics The Netherlands (1991–93) Prosp3 Recruitment not reported Popl: 250 patients/ index: 250 shoulders/ ref. test: 41[m] shoulders *Mean age*: 20–75 *Sex*: 41% male *Outcomes*: Any; FT; PT *Prevalence*: 49; 12; 37%	Patients referred for shoulder US with signs and symptoms suggestive of RC impingement or tear, biceps tendinitis, labrum and capsular abnormalities or nonspecific shoulder pain [m] 81 underwent US and ref. test but first 40 were used to learn the technique and remaining 41 included in study	*Scanner*: 7.5-MHz linear transducer *Technique*: Used combination of techniques as described by van Holsbeek,1992 and Hedtmann, 1988. Included a dynamic examination	*Ref. test*: Arthroscopy/surgery/arthrography. Conventional arthrography or CTA (if no FT tear was present on conventional arthrography), surgery and/or arthroscopy. No further details provided. Several patients underwent more than one ref. test. 34 underwent some form of surgery, 1 received conventional arthrography and 6 underwent CTA. *Test interval*: Arthrography performed less than 1 week after US, surgical procedures performed within 6 months of US *Further investigations*: Not reported *Diagnostic criteria*: Criteria used were provided (no reference). Normal: RC homogeneous, echogenic arc from bursa, peribursal fat and deltoid. Non-homogeneous: focal irregularity with hyper- and hypoechogenic foci <2 mm. PT: hypoechoic zone within RC ⩾3 mm, thinning of RC. FT: hyperechoic zone through cuff or non-visualisation of RC. Not stated whether non-homogeneous were regarded as normal or PT RCT *Test interpreters*: Not reported. First 40 patients used to learn technique *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Vick, 1990[74] Radiology Italy (years not reported) Cannot tell design Consecutive Popl: 115 patients/index: 115 shoulders/ref. test: 81 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: FT *Prevalence*: 30%	Patients with clinically suspected RCT who had US and arthrography/surgery	*Scanner*: 5-MHz linear phased-array transducer *Technique*: Technique used was based on a review of the literature (Middleton, 1984,[122] 1985[48]; Mack, 1985[46]; Crass, 1987[127]) and experience gained from scanning normal subjects	*Ref. test*: Arthrography (79)/surgery (2). Details provided *Test interval*: Same day for US and arthrography. Not reported for US and surgery *Further investigations*: Not reported *Diagnostic criteria*: Criteria used to detect FT RCT described (Mack, 1985). Required focal defect in RC or complete absence or non-visualisation of RC *Test interpreters*: Single radiologist – experience relatively limited? *Ref. test interpretation*: Radiologist, number not reported
Wallny, 2001[42] Orthopaedics Germany (years not reported) Prosp2 Recruitment not reported Popl: NR patients/ index: 40 shoulders/ ref. test: 40 shoulders *Mean age*: 54 (38–79) *Sex*: 62% male *Outcomes*: Any *Prevalence*: 58%	Patients with shoulder pain with histories and physical examination suggestive of RC lesions admitted for surgical exploration/or repair Indications for surgery based on results of clinical assessment and MRI scan results. Excluded if prior shoulder surgery or previous fracture of humeral head	Compared 2D and 3D US *Scanner*: 10-MHz transducer *Technique*: standard 2D (numerous details of technology). The region of interest needed to be defined by 2D US before 3D could be undertaken	*Ref. test*: Arthroscopy/surgery. Diagnostic and/or therapeutic arthroscopy (23), open surgery (17). No details provided *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria reported. Same for 2D and 3D US. (1) Marked thinning, sudden changes of calibre and total absence of the cuff were defined as formal 'external' criteria (Bachmann, 1997[141]). (2) Hyper- and/or hypoechoic zones defined as 'structural' criteria (Bachmann, 1997[141]; Hedtmann, 1995[142]; Pattee, 1988[69]). A partial rupture defined as no more than loss of 1/4 to 1/2 of FT intact RC (Fukuda, 1996).[143] *Test interpreters*: Not reported. 2D US done immediately prior and by same person as 3D scan so results not independent *Ref. test interpretation*: Orthopaedic surgeon with long-standing experience of shoulder surgery

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Wiener, 1993[56] Radiology/orthopaedics USA (1985–92) Cannot tell design Recruitment not reported Popl: 800 patients/ index: 225 shoulders/ ref. test: 225 shoulders *Mean age*: 59 (21–81) *Sex*: 47% male *Outcomes*: Any; FT; PT *Prevalence*: 70; 40; 30%	Patients with signs and symptoms referable to impingement and suspected RCTs who were referred for US and ultimately required surgical management Sample biased in favour of those with persistent signs and symptoms or functional limitations of the shoulder joint (surgical proof lacking for 575)	*Scanner*: 7.5-MHz transducer with additional 5-MHz linear array used for large shoulders *Technique*: Axial and sagittal planes, arm in neutral and in internal rotation. Technique fully described (Crass, 1988[144])	*Ref. test*: Surgery – arthroscopic or open according to sonographic classification. Patients with intact normal or partial tears who did not respond to medical management within a designated time underwent arthroscopic decompression of the subacromial space (those who did respond were not included in the study sample). For small full RCT arthroscopic decompression plus limited splitting of deltoid. Large/massive tears had formal open surgical exposure for examination and repair of the RC *Test interval*: Not reported *Further investigations*: Not stated *Diagnostic criteria*: Criteria for classification provided (presumably used those of Crass, 1988[144]). PT tear: focal hyperechoic zone within RC, small hyperechoic disturbances of internal/external surfaces of cuff, or dominant linear echogenic focus within cuff. FT tear: hyperechoic zone extending through entire thickness of RC, or segmental or complete loss of RC substance with visualisation of tear margins *Test interpreters*: Not reported *Ref. test interpretation*: No details. Not blinded – may not have looked too hard for tears in these those with negative US

Appendix 7

Ultrasound: sensitivity and specificity results

Study	Comparison	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Ahovuo, 1989[73]		–	–	0.75 (0.57 to 0.87)	0.95 (0.86 to 0.98)	–	–
Arslan, 1999[50]	T+: biceps effusion	0.35 (0.24 to 0.49)	0.74 (0.61 to 0.84)	–	–	–	–
	T+: bursal fluid	0.8 (0.3 to 0.19)	0.94 (0.85 to 0.98)	–	–	–	–
	T+: both present	0.12 (0.6 to 0.23)	0.91 (0.80 to 0.96)	–	–	–	–
	T+: either present[a]	0.55 (0.41 to 0.68)	0.59 (0.46 to 0.71)	–	–	–	–
Brandt, 1989[70]	Prosp interp[a]	0.75 (0.57 to 0.87)	0.43 (0.27 to 0.61)	–	–	–	–
	Retrosp interp	0.68 (0.49 to 0.82)	0.90 (0.74 to 0.97)	–	–	–	–
Brenneke, 1992[51]		0.78 (0.67 to 0.87)	0.82 (0.70 to 0.90)	0.95 (0.83 to 0.99)	0.93 (0.85 to 0.97)	0.41 (0.25 to 0.59)	0.91 (0.84 to 0.96)
Burk, 1989[71]		0.60 (0.36 to 0.80)	0.57 (0.37 to 0.74)	–	–	–	–
Chiou, 1999[76]		0.98 (0.87 to 1.00)	0.87 (0.62 to 0.96)	–	–	–	–
Crass, 1988[40]	Ind excl	0.90 (0.79 to 0.95)	0.92 (0.81 to 0.97)	–	–	–	–
	Ind T+	0.91 (0.82 to 0.96)	0.79 (0.67 to 0.88)	–	–	–	–
	Ind T–[a]	0.79 (0.67 to 0.87)	0.93 (0.84 to 0.97)	–	–	–	–
De Muynck 1994[57]		0.81 (0.63 to 0.92)	0.82 (0.52 to 0.95)	–	–	–	–

continued

Study	Comparison	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Drakeford, 1990[68]		0.92 (0.65 to 0.99)	0.95 (0.83 to 0.99)	–	–	–	–
Farin, 1990[52]	S I–III	0.81 (0.66 to 0.91)	0.95 (0.87 to 0.98)	–	–	–	–
	S I–II[a]	0.71 (0.50 to 0.86)	0.96 (0.90 to 0.99)	–	–	–	–
Gratz, 1998[43]	Ind T–[a]	0.83 (0.44 to 0.97)	0.91 (0.62 to 0.98)	–	–	–	–
	Ind T+	1.00 (0.61 to 1.00)	0.55 (0.28 to 0.79)	–	–	–	–
	Ind T–/+	1.00 (0.61 to 1.00)	0.73 (0.43 to 0.90)	–	–	–	–
Hodler, 1988[79]	Surgery[a]	1.00 (0.90 to 1.00)	0.75 (0.51 to 0.90)	–	–	–	–
	Arthrog	0.93 (0.77 to 0.98)	0.82 (0.52 to 0.95)	–	–	–	–
Hodler, 1991[59]		–	–	0.93 (0.70 to 0.99)	0.78 (0.45 to 0.94)		
Kurol, 1991[44]	T+: echog	0.54 (0.35 to 0.72)	0.82 (0.66 to 0.92)	–	–	–	–
	T+: thinning	0.42 (0.24 to 0.61)	0.88 (0.73 to 0.95)	–	–	–	–
	T+: either[a]	0.67 (0.47 to 0.82)	0.74 (0.57 to 0.85)	–	–	–	–
Mack, 1988[47]		0.88 (0.76 to 0.94)	0.96 (0.87 to 0.99)	–	–	–	–
Martin–Hervas, 2001[66]		0.71 (0.54 to 0.83)	0.67 (0.48 to 0.81)	0.58 (0.39 to 0.74)	1.00 (0.90 to 1.00)	0.13 (0.2 to 0.47)	0.68 (0.55 to 0.79)

continued

Study	Comparison	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Middleton, 1986[41]	Ind excl	0.91 (0.77 to 0.97)	0.91 (0.82 to 0.96)	–	–	–	–
	Ind T+	0.92 (0.78 to 0.97)	0.86 (0.76 to 0.92)	–	–	–	–
	Ind T–[a]	0.86 (0.71 to 0.94)	0.92 (0.84 to 0.96)	–	–	–	–
Miller, 1989[72]		0.58 (0.39 to 0.74)	0.93 (0.79 to 0.98)	–	–	–	–
Misamore, 1991[58]		0.33 (0.19 to 0.52)	0.60 (0.23 to 0.88)	–	–	–	–
Naredo, 1999[60]		–	–	0.89 (0.67 to 0.97)	1.00 (0.82 to 1.00)	0.92 (0.67 to 0.99)	0.91 (0.73 to 0.98)
Nelson, 1991[64]		–	–	0.60 (0.23 to 0.88)	0.93 (0.69 to 0.99)	0.36 (0.15 to 0.65)	0.75 (0.41 to 0.93)
Olive, 1992[45]		–	–	0.90 (0.74 to 0.96)	0.91 (0.78 to 0.96)	–	–
Paavolainen, 1994[61]		–	–	0.74 (0.55 to 0.87)	0.95 (0.78 to 0.99)	–	–
Pattee, 1988[69]		0.77 (0.61 to 0.88)	0.65 (0.41 to 0.83)	–	–	–	–
Read, 1998[36]		0.79[1] (0.63 to 0.90)	0.88[1] (0.53 to 0.98)	1.00 (0.72 to 1.00)	0.97 (0.84 to 0.99)	0.46 (0.23 to 0.71)	0.97 (0.83 to 0.99)
Roberts, 2001[77]		0.76 (0.53 to 0.90)	1.00 (0.65 to 1.00)	0.80 (0.49 to 0.94)	1.00 (0.78 to 1.00)	0.71 (0.36 to 0.92)	1.00 (0.82 to 1.00)
Shipley, 1985[65]		0.70 (0.40 to 0.89)	1.00 (0.34 to 1.00)	–	–	–	–

continued

Study	Comparison	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Soble, 1989[75]		0.92 (0.76 to 0.98)	0.84 (0.71 to 0.91)	–	–	–	–
Sonnabend, 1997[53]		0.70 (0.58 to 0.79)	0.95 (0.84 to 0.99)	0.84 (0.71 to 0.92)	0.92 (0.83 to 0.97)	0.25 (0.12 to 0.45)	0.99 (0.94 to 1.00)
Swen, 1998[62]		–	–	0.85 (0.58 to 0.96)	1.00 (0.68 to 1.00)	–	–
Swen, 1999[63]	R1[a]	–	–	0.92 (0.67 to 0.99)	0.88 (0.53 to 0.98)	–	–
	R2	–	–	0.69 (0.42 to 0.87)	0.88 (0.53 to 0.98)	–	–
Takagishi, 1996[54]		0.72 (0.60 to 0.82)	0.89 (0.79 to 0.95)	0.76 (0.61 to 0.87)	1.00 (0.96 to 1.00)	0.50 (0.30 to 0.70)	0.90 (0.83 to 0.95)
Teefey, 2000[55]		0.90 (0.81 to 0.95)	0.85 (0.64 to 0.95)	1.00 (0.94 to 1.00)	0.91 (0.78 to 0.97)	0.47 (0.25 to 0.70)	0.93 (0.85 to 0.97)
van Holsbeeck, 1995[78]		0.98 (0.89 to 1.00)	0.83 (0.61 to 0.94)	1.00 (0.90 to 1.00)	1.00 (0.90 to 1.00)	0.93 (0.70 to 0.99)	0.94 (0.84 to 0.98)
Van Moppes, 1995[67]		0.80 (0.58 to 0.92)	0.95 (0.77 to 0.99)	1.00 (0.57 to 1.00)	1.00 (0.90 to 1.00)	0.73 (0.48 to 0.89)	0.96 (0.81 to 0.99)
Vick, 1990[74]		–	–	0.67 (0.47 to 0.82)	0.93 (0.83 to 0.97)	–	–
Wallny, 2001[42]	2D[a]	0.74 (0.54 to 0.87)	0.82 (0.59 to 0.94)	–	–	–	–
	3D	0.91 (0.73 to 0.98)	0.82 (0.59 to 0.94)	–	–	–	–
Wiener, 1993[56]		0.95 (0.90 to 0.97)	0.94 (0.86 to 0.98)	0.93 (0.86 to 0.97)	0.99 (0.96 to 1.00)	0.94 (0.86 to 0.98)	0.97 (0.93 to 0.99)

Se, sensitivity; Sp, specificity; T+, criteria for test positive; T–, criteria for test negative; Prosp, prospective; Retrosp, retrospective; S, stages of impingement syndrome; echog, echogenicity; Ind, indeterminate results; R, reader; D, dimensional.
[a] Indicates set of results used for meta-analysis.

Appendix 8

Ultrasound: likelihood ratio results

Study	Comparison	Any tear		Full tear		Partial tear	
		LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)
Ahovuo, 1989[73]		–	–	15.0 (4.9 to 46.1)	0.3 (0.1 to 0.5)	–	–
Arslan, 1999[50]	T+: biceps effusion	1.4 (0.8 to 2.4)	0.9 (0.7 to 1.1)	–	–	–	–
	T+: bursal fluid	1.4 (0.3 to 6.0)	1.0 (0.9 to 1.1)	–	–	–	–
	T+: both present	1.3 (0.4 to 3.9)	1.0 (0.9 to 1.1)	–	–	–	–
	T+: either present[a]	1.3 (0.9 to 2.0)	0.8 (0.5 to 1.1)	–	–	–	–
Brandt, 1989[70]	Prosp interp[a]	1.3 (0.9 to 1.9)	0.6 (0.3 to 1.2)	–	–	–	–
	Retrosp interp	6.8 (2.3 to 20.5)	0.4 (0.2 to 0.6)				
Brenneke, 1992[51]		4.3 (2.4 to 7.7)	0.3 (0.2 to 0.4)	12.9 (6.0 to 28.1)	0.1 (0.0 to 0.2)	4.7 (2.1 to 10.6)	0.6 (0.5 to 0.9)
Burk, 1989[71]		1.4 (0.7 to 2.6)	0.7 (0.3 to 1.4)	–	–	–	–
Chiou, 1999[76]		7.3 (2.0 to 26.6)	0.0 (0.0 to 0.2)	–	–	–	–
Crass, 1988[40]	Ind excl	11.2 (4.4 to 8.8)	0.1 (0.1 to 0.2)	–	–	–	–
	Ind T+	4.4 (2.6 to 7.3)	0.1 (0.1 to 0.2)	–	–	–	–
	Ind T–[a]	11.4 (4.4 to 9.7)	0.2 (0.1 to 0.4)	–	–	–	–
De Muynck 1994[57]		4.5 (1.3 to 15.9)	0.2 (0.1 to 0.5)	–	–	–	–
Drakeford, 1990[68]		17.4 (4.5 to 67.8)	0.1 (0.0 to 0.6)	–	–	–	–
Farin, 1990[52]	S I–III	17.6 (5.8 to 53.6)	0.2 (0.1 to 0.4)	–	–	–	–
	S I–II[a]	19.3 (6.2 to 60.5)	0.3 (0.2 to 0.6)	–	–	–	
Gratz, 1998[43]	Ind T–[a]	9.2 (1.4 to 61.5)	0.2 (0.0 to 1.1)	–	–	–	–
	Ind T+	2.2 (1.2 to 4.2)	0.1 (0.0 to 2.0)	–	–	–	–
	Ind T–/+	3.7 (1.4 to 9.6)	0.1 (0.0 to 1.5)	–	–	–	–
Hodler, 1988[79]	Surgery (n = 51)	4.0 (1.7 to 9.3)	0.1 (0.0 to 0.3)	–	–	–	–
	Arthrog (n = 39)	5.1 (1.5 to 18.0)	0.1 (0.0 to 0.3)		–	–	–

continued

Study	Comparison	Any tear		Full tear		Partial tear	
		LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)
Hodler, 1991[59]		–	–	4.2 (1.2 to 14.4)	0.1 (0.0 to 0.6)	–	–
Kurol, 1991[44]	T+: echog	3.1 (1.4 to 6.9)	0.6 (0.4 to 0.9)	–	–	–	–
	T+: thinning	3.5 (1.3 to 10.0)	0.7 (0.5 to 0.9)	–	–	–	–
	T+: either[a]	2.5 (1.3 to 4.7)	0.5 (0.2 to 0.8)	–	–	–	–
Mack, 1988[47]		21.9 (5.6 to 85.6)	0.1 (0.1 to 0.3)	–	–	–	–
Martin–Hervas to 2001[66]		2.1 (1.2 to 3.8)	0.4 (0.2 to 0.8)	41.3 (2.6 to 660.7)	0.4 (0.3 to 0.7)	1.3 (0.9 to 1.8)	0.4 (0.1 to 2.5)
Middleton, 1986[41]	Ind excl	10.0 (4.6 to 21.7)	0.1 (0.0 to 0.3)	–	–	–	–
	Ind T+	6.4 (3.6 to 11.5)	0.1 (0.0 to 0.3)	–	–	–	–
	Ind T–[a]	10.8 (4.9 to 23.4)	0.2 (0.1 to 0.3)	–	–	–	–
Miller, 1989[72]		8.7 (2.2 to 34.4)	0.5 (0.3 to 0.7)	–	–	–	–
Misamore, 1991[58]		0.8 (0.3 to 2.8)	1.1 (0.5 to 2.4)	–	–	–	–
Naredo, 1999[60]		–	–	33 (2.1 to 511.4)	0.1 (0.0 to 0.4)	10.6 (2.8 to 40.3)	0.1 (0.0 to 0.6)
Nelson, 1991[64]		–	–	8.4 (1.1 to 63.3)	0.4 (0.1 to 1.3)	1.5 (0.3 to 6.1)	0.8 (0.5 to 1.5)
Olive, 1992[45]		–	–	9.6 (3.8 to 24.7)	0.1 (0.0 to 0.3)	–	–
Paavolainen, 1994[61]		–	–	16.3 (2.4 to 112.0)	0.3 (0.1 to 0.5)	–	–
Pattee, 1988[69]		2.2 (1.1 to 4.3)	0.4 (0.2 to 0.7)	–	–	–	–
Read, 1998[36]		6.4 (1.0 to 40.1)[l]	0.2 (0.1 to 0.5)[l]	32.0 (4.6 to 220.3)	0.0 (0.0 to 0.7)	13.4 (1.8 to 100.2)	0.6 (0.3 to 0.9)
Roberts, 2001[77]		12 (0.8 to 178.0)	0.2 (0.1 to 0.6)	23.2 (1.5 to 360.5)	0.2 (0.1 to 0.7)	24.7 (1.5 to 396.0)	0.3 (0.1 to 0.9)
Shipley, 1985[65]		4.1 (0.3 to 53.0)	0.3 (0.1 to 0.8)	–	–	–	–
Soble, 1989[75]		5.7 (3.0 to 10.8)	0.1 (0.0 to 0.4)	–	–	–	–

continued

Study	Comparison	Any tear		Full tear		Partial tear	
		LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)
Sonnabend, 1997[53]		14.3 (3.7 to 55.6)	0.3 (0.2 to 0.5)	11.0 (4.7 to 25.7)	0.2 (0.1 to 0.3)	21.5 (2.7 to 170.1)	0.8 (0.6 to 1.0)
Swen, 1998[62]		–	–	7.5 (2.5 to 22.0)	0.2 (0.1 to 0.4)	–	–
Swen, 1999[63]	R1[a]	–	–	7.4 (1.2 to 46.5)	0.1 (0.0 to 0.6)	–	–
	R2	–	–	5.5 (0.9 to 35.9)	0.4 (0.1 to 0.8)	–	–
Takagishi, 1996[54]		6.6 (3.2 to 13.6)	0.3 (0.2 to 0.5)	128.6 (8.1 to 2050.8)	0.2 (0.1 to 0.4)	5.1 (2.4 to 10.6)	0.6 (0.4 to 0.9)
Teefey, 2000[55]		6.0 (2.1 to 17.1)	0.1 (0.1 to 0.2)	11.7 (4.0 to 34.4)	0.0 (0.0 to 0.1)	6.6 (2.6 to 17.0)	0.6 (0.4 to 0.9)
van Holsbeeck, 1995[78]		5.9 (2.1 to 16.5)	0.0 (0.0 to 0.2)	67.0 (4.3 to 1050.2)	0.0 (0.0 to 0.2)	16.2 (5.3 to 48.9)	0.1 (0.0 to 0.5)
Van Moppes, 1995[67]		16.8 (2.5 to 15.2)	0.2 (0.1 to 0.5)	67.8 (4.3 to 1075.6)	0.1 (0.0 to 1.2)	19.1 (2.7 to 133.5)	0.3 (0.1 to 0.6)
Vick, 1990[74]		–	–	9.5 (3.5 to 25.5)	0.4 (0.2 to 0.6)	–	–
Wallny, 2001[42]	2D[a]	4.2 (1.5 to 12.0)	0.3 (0.2 to 0.7)	–	–	–	–
	3D	5.2 (1.8 to 14.6)	0.1 (0.0 to 0.4)	–			
Wiener, 1993[56]		15.9 (6.1 to 41.2)	0.1 (0.0 to 0.1)	126.0 (17.9 to 888.7)	0.1 (0.0 to 0.1)	29.6 (12.4 to 70.2)	0.1 (0.0 to 0.2)

LR+, positive likelihood ratio; LR–, negative likelihood ratio; T+, criteria for test positive; T–, criteria for test negative; Prosp, prospective; Retrosp, retrospective; S, stages of impingement syndrome; echog, echogenicity; Ind, indeterminate results; R, reader; D, dimensional.
[a] Indicates set of results used for metaanalysis.

Appendix 9

MRI: detailed study methods

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Balich, 1997[83] Diagnostic radiology, USA (1989–95) Retrosp Recruitment not reported Popl: NR patients/ index: 222 shoulders/ ref. test: 222 shoulders *Mean age*: 45 (16–76) *Sex*: 58% male *Outcomes*: Any; FT; PT *Prevalence*: 3, 20, 18%	Symptomatic patients who had undergone both MRI and arthroscopy Excluded if interval between index and ref. test >365 days	*MRI unit*: 1.5 T; dedicated angled surface coil *Sequences*: Both conventional and fast spin-echo techniques with fat suppression[a] *Planes*: OC, oblique parasagittal (OP), and AX [a] Over time, the imaging parameters changed to include the use of a fast spin-echo sequence and a decrease in the field of view	*Ref. test*: Arthroscopy. Note area investigated differed across patients: glenohumeral and subacromial joint (141); subacromial bursa (63); glenohumeral joint space alone (18) *Test interval*: Mean 66 (0–364) days *Further investigations*: No details *Diagnostic criteria*: Discussion states 'no diagnostic criteria were supplied to the readers before the study' *Test interpreters*: Five readers, in random order, independently twice. Third-year resident with 2 months musculoskeletal radiology rotations; angiography/cross-sectional imaging fellow; 3 musculoskeletal radiologists incl. academic radiologist and 2 private radiologists with up to 7 years experience Readers first blinded to all but patients age and sex and then a second review was immediately performed with the arthroscopy report in hand. Data from blind reading only used here *Ref. test interpretation*: Typewritten arthroscopy report used. Not stated whether results of MRI known at time of arthroscopy nor what information directed arthroscopy examination
Birtane, 2001[88] Physical ther. & rehab. Turkey (years not reported) Prosp1 Consecutive Popl: 120 patients/index: 120 shoulders/ref. test: 120 shoulders (125 shoulders) *Mean age*: 51.6 (SD 13.9) *Sex*: 40% male *Outcomes*: IS *Prevalence*: 70%	Patients with shoulder pain Excluded if inflammatory or systemic diseases with shoulder involvement; acute traumatic conditions; post-operative conditions; neck or elbow disorders Patients referred from rheumatology or applied directly	Compared clinical examination and MRI (clin. exam. reported in Calis *et al*.;[30] Appendix 5) *MRI unit*: 1.0 T; shoulder coil *Sequences*: Fast and gradient spin-echo T1 and T2 sequences *Planes*: OC	*Ref. test*: Subacromial injection test – details provided *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Zlatkin stages of SIS. Stage 1 and above accepted as MRI positive *Test interpreters*: Experienced radiologist *Ref. test interpretation*: Not reported – states experienced hands (Calis *et al*. paper[30])

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Blanchard, 1999[98] Radiology/orthopaedics UK (years not reported) Prosp3 Other Popl: 117 patients/ index: 104 shoulders/ ref. test: 38[a] shoulders *Mean age*: M: 50.6 (24–74) F: 49.7 (28–73) *Sex*: 55% male; *n* = 57 *Outcomes*: FT *Prevalence*: 30%	Patients with shoulder problems referred to a specialist orthopaedic shoulder clinic for possible surgery [a] All subjects followed up over 6 months but only those undergoing surgery used to estimate accuracy	*MRI unit*: 0.5 or 1.5 T; surface coils *Sequences*: 1. Dual-echo images (providing proton-density and T2-weighted images); 2. T1-weighted images *Planes*: 1. OC and sagittal; 2. AX Arthrography also evaluated (data not extracted – sensitivity found to be lower than MRI but specificity higher)	*Ref. test*: Arthroscopy/surgery. Options either open surgery, arthroscopic surgery or discharge/referral to another clinician. No further details *Test interval*: Up to 6 months (reviewed retrospectively). Authors note that patients suitable for arthroscopic surgery treated more promptly *Further investigations*: All patients also underwent arthrography as index test. Results did not contribute to reference standard diagnosis *Diagnostic criteria*: Based on standard criteria in literature (Iannotti, 1991[17]) *Test interpreters*: Radiologist (*n* = 2) with over 3 years experience of shoulder MRI *Ref. test interpretation*: Orthopaedic surgeon NB: main aims of study: diagnostic impact of index tests (diagnostic confidence) and therapeutic impact (changes in management)
Burk, 1989[71] Radiology/orthopaedics USA (years not reported) Prosp1 Recruitment not reported Popl: 41 patients/index: 38 shoulders MRI; 23 US/ ref. test: 38 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 58%	Patients referred for evaluation of possible RCTs 3 excluded owing to claustrophobia	*MRI unit*: 1.5 T; loop gap resonator surface coil *Sequences*: Spin-echo (TR/TE: 600/20, 2000/40, and 600/25) *Planes*: AX, OC and sagittal US also evaluated (see Appendix 6)	*Ref. test*: Double-contrast arthrography. Details provided *Test interval*: Sequentially and same day *Further investigations*: US in 23 patients (5 MHz) and surgery in 16 (no details of surgery). Not clear whether surgical results contributed to reference diagnosis *Diagnostic criteria*: Criteria used to diagnose tear were reported (no reference given) *Test interpreters*: Musculoskeletal radiologist with extensive experience in MR imaging but limited experience in the diagnosis of RCTs by MRI *Ref. test interpretation*: Experienced musculoskeletal radiologist who was not involved in the interpretation of the MR images

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Evancho, 1988[89] Radiology/orthopaedics USA (1986–7) Cannot tell design Recruitment not reported Popl: 44 patients/ index: 31[b] shoulders/ ref. test: 31 shoulders *Mean age*: Range 25–77 *Sex*: Not reported *Outcomes*: Any; FT; PT *Prevalence*: 42; 36; 14%	Symptomatic patients with suspected RCTs [b] 8 excluded for significant motion artifacts on MRI. 5 excluded as results of ref. test not available	*MRI unit*: 0.5 or 1.5 T *Sequences*: Spin-echo *Planes*: Oblique	*Ref. test*: Arthroscopy (n = 19)/arthrography (n = 12). Details provided but criteria for choice of ref. test not provided *Test interval*: Not described *Further investigations*: Not reported *Diagnostic criteria*: Criteria for RCT provided (no references) *Test interpreters*: Not reported. MRI interpreted prospectively without knowledge of the diagnostic test results or surgery. Then reviewed retrospectively after arthroscopic results known *Ref. test interpretation*: Initial interpretation of arthrography used for comparison with MR. Procedure was videotaped and retrospectively reviewed by an experienced orthopaedic surgeon
Hodler, 1991[59] Radiology US (years not reported) Prosp2 Recruitment not reported Popl: 23 patients (24 shoulders)/index: 24 shoulders/ref. test: 24 shoulders *Mean age*: 57.8 (39–84) *Sex*: 43% male *Outcomes*: FT *Prevalence*: 62%	Patients with chronic shoulder pain and suspected RCT Required pathogenesis other than degenerative disease: rheumatoid arthritis (11), chronic pain after trauma (3), calcifying tendinitis (3) and degenerative changes (6)	Compared MRI and US (see Appendix 6) *MRI unit*: 1.5 T; flexible receive-only surface coil *Sequences*: T1- and T2-weighted spin-echo *Planes*: OC NB: hardware and software upgraded during study, improving image quality	*Ref. test*: Double-contrast arthrography. Details provided *Test interval*: Not reported *Further investigations*: None reported *Diagnostic criteria*: Not clearly reported and no criteria provided *Test interpreters*: Radiologists experienced in MRI (n = 4). Blinded to US results. No further details

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Hodler, 1992[90] Radiology/orthopaedics USA (years not reported) Cannot tell design Recruitment not reported Popl: 150 patients/ index: 36 shoulders/ ref. test: 36 shoulders *Mean age*: 42.5 (17–69) *Sex*: 67% male *Outcomes*: Any; FT; PT *Prevalence*: 57; 11; 47%	Patients who had MRI and whose arthroscopy reports were available	Compared MRI and MRA (see Appendix 12) *MRI unit*: 1.5 T *Sequences*: Proton-density, T2-weighted *Planes*: OC States no overlap with 1991 study[59]	*Ref. test*: Arthroscopy: details provided. Examined glenohumeral and subacromial bursa (n = 32). 3 young patients with shoulder instability and 1 with adhesion capsulitis did not undergo bursal examination *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria provided (no reference given). Superficial fraying included in 'no tear' category *Test interpreters*: MRA alone and standard MRI images read by 3 experienced osteoradiologists (n = 3); read without knowledge of original report or clinical data. Consensus reading sought *Ref. test interpretation*: Performed by experienced surgeons. The detailed arthroscopic reports were used for this study
Iannotti, 1991[17] Orthopaedics/radiology US (years not reported) Retrosp Recruitment not reported Popl: 127 shoulders/ index: 112[c] shoulders/ ref. test: 73 shoulders (further 15 volunteers who did not get surgery included in analyses) *Mean age*: 40 (estimated) *Sex*: Not reported *Outcomes*: FT; IS *Prevalence*: 37; ?%	Patients undergoing surgery for lesions of the RC or glenohumeral capsule and glenoid labrum plus asymptomatic volunteers Symptoms relating to RC or labral lesions with no history of dislocation [c] 39 who had capsule-labrum diagnoses at surgery excluded from analyses	*MRI unit*: 1.5 T; receive-only surface coils *Sequences*: 1. T1-weighted and proton density; 2. T2-weighted *Planes*: 1. AX, OS and OC; 2. OC	*Ref. test*: Arthroscopy or open surgery. Area investigated varied between patients (73 RC only; 39 glenohumeral capsule glenoid labrum). Former included in analyses along with the asymptomatic volunteers (total n = 88). *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis described (Zlatkin, 1989[16]). *Test interpreters*: Radiologists: 3 interpreted the images for the 32 full tears and the images from the asymptomatic volunteers. The remainder interpreted by only one radiologist (implies knowledge of MRI result)

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Jaovisidha, 1999[91] Radiology US (years not reported) Prosp1 Consecutive Popl: 32[d] patients/index: 27 shoulders/ref. test: 8 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any tear *Prevalence*: 25%	Consecutive patients undergoing MRI to evaluate RC [d] 5 patients could not perform exercise	Compared MRI before and after active or passive exercise *MRI unit*: 1.5 T; shoulder surface coil *Sequences*: Fast spin-echo T2-weighted *Planes*: OC	*Ref. test*: Arthroscopy *Test interval*: Not reported. 26-month follow-up period *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis provided (with various references, including Zlatkin 1989,[16] Rafii,[145] Iannotti,[17] Kneeland[95]) *Test interpreters*: 2 musculoskeletal radiologists blinded to clinical information. Consensus decision used. Post-exercise images interpreted alongside pre-exercise images
Kieft, 1988[113] Diagnostic radiology/ orthopaedics Netherlands (years not reported) Prosp2 Recruitment not reported Popl: NR patients/index: 10 shoulders/ref. test: 10 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 30%	Clinically suspected RC impingement	*MRI unit*: 0.5 T; surface coil. *Sequences*: Spin-echo with T1 and T2 weighting. *Planes*: Transverse, coronal, sagittal and oblique planes, perpendicular and parallel to the glenoid surface	*Ref. test*: Double-contrast arthrography. Details provided *Test interval*: Not described *Further investigations*: Surgical treatment with anterior acromioplasty performed in 6 patients *Diagnostic criteria*: Criteria for RCT provided (no references) *Test interpreters*: Radiologists ($n = 2$) with access to clinical history *Ref. test interpretation*: Not reported
Kneeland, 1987[95] Radiology/orthopaedics USA (years not reported) Retrosp Recruitment not reported Popl: NR patients/index: 25 shoulders (26 shoulders)/ ref. test: 26 shoulders *Mean age*: range 24–67 *Sex*: 60% male *Outcomes*: Any *Prevalence*: 85%	Patients with known or suspected RCTs based on clinical signs	*MRI unit*: 1.5 T; shoulder coil *Sequences*: Spin-echo pulse sequences with proton density-weighting and T2-weighting *Planes*: Coronal and sagittal	*Ref. test*: Arthrography in 24 shoulders, surgery in 2. No details *Test interval*: Within 30 days *Further investigations*: Not reported *Diagnostic criteria*: Criteria reported (no reference given). Tear reported if discontinuity of the cuff could be identified and increased signal intensity *Test interpreters*: MR and arthrography reviewed in conference by 3 authors. Tests results determined by consensus *Ref. test interpretation*: As above

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Martin-Hervas, 2001[66] Orthopaedics Spain (1998) Prosp3 Consecutive Popl: 140 shoulders/index: 140 US, 61 MRI[e]/ref. test: 72 *Mean age*: Not reported *Sex*: [e]41% male *Outcomes*: Any; FT; PT *Prevalence*: 56; 43; 13%	Patients with shoulder pain and limited movement fulfilled criteria for orthopaedic surgery protocol (n = 72) Only those undergoing US and MRI included in accuracy estimates (n = 61) [e] 11 unable to have MRI owing to claustrophobia, metal implants or pacemaker	Compared ultrasound and MRI, alone and in combination (see Appendix 6 for US details) *MRI unit*: 0.5 T; surface coil *Sequences*: 1. Spin-echo T1-weighted images; 2. gradient T2 sequences *Planes*: 1. coronal and oblique; 2. axial and oblique	*Ref. test*: Arthroscopy (40), open surgery (10) or both (11) *Test interval*: <6 months in all cases *Further investigations*: Not reported *Diagnostic criteria*: Criteria provided (no references) *Test interpreters*: Musculoskeletal radiologists – no further details *Ref. test interpretation*: Not described
Morrison, 1990[94] Sports medicine centre USA (1987–8) Prosp1 Consecutive Popl: 106[f] patients/ index: 100 shoulders/ ref. test: 100 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any; FT *Prevalence*: 55; 51%	Patients (athletes) with symptoms of subacromial impingement for >6 months Most had previous treatment with NSAIDs and local steroid injections [f] 4 patients too large for MR scanner; 2 suffered claustrophobia	*MRI unit*: 1.5 T; 3-in angled pair shoulder surface coil *Sequences*: Intermediate and T2-weighted *Planes*: True sagittal plane NB: early in the investigation it became apparent that comparison of intermediate and T2-weighted images could identify RCTs better	*Ref. test*: Single-contrast exercise arthrography. Details provided *Test interval*: Not stated *Further investigations*: All arthrography test positives had open operative repair. Some arthrography test negatives had arthroscopic subacromial decompression *Diagnostic criteria*: Criteria reported, no reference given *Test interpreters*: All MRIs performed by one radiologist. Arthrogram and MR interpreted separately without knowledge of patient's name or results of the companion study *Ref. test interpretation*: All arthrograms were performed by different radiologists. See above

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Nelson, 1991[64] Radiology/orthopaedics US (years not reported) Prosp1 Recruitment not reported Popl: 21 patients/ index: 21 MRI, 19 US/ ref. test: 21 shoulders *Mean age*: 42 (18–70) *Sex*: 76% male *Outcomes*: FT; PT *Prevalence*: 33; 57%	Patients with shoulder pain resistant to non-operative treatment History of shoulder pain >3 months before first consultation	Compared MRI and US (see Appendix 6 for US details) *MRI unit*: 1.5 T; Helmholtz surface shoulder coil *Sequences*: 1. Spin-echo T1-weighted localising images; 2. contiguous T1-weighted proton density and T2-weighted *Planes*: 1. Axial; 2. coronal	*Ref. test*: Arthroscopy (n = 20) or open operative procedure (n = 3). *Test interval*: Not reported *Further investigations*: Two patients underwent arthroscopy and open surgery, one in whom avascular necrosis of the head of the humerus after fracture of neck of humerus and the other who fell after arthroscopy and needed an arthroscopic Bankart repair *Diagnostic criteria*: Criteria for diagnosis provided (Zlatkin, 1989[16]) *Test interpreters*: Radiologists (n = 2). Complete set of studies for each patient was read by one radiologist (without clinical information), whose report influenced choice of treatment. These findings were used as final reading for the purpose of the study. All studies were read by second radiologist, who was unaware of the results of the other studies, type of clinical management and operative findings *Ref. test interpretation*: As above
Patten, 1994[92] Radiology USA (years not reported) Retrosp Consecutive Popl: 50 patients/ index: 50 shoulders/ ref. test: 25[g] shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 40%	Not reported [g] Authors presumed 25 patients who did not get ref test were test negatives based on clinical findings only – these have been included in accuracy results	MRI: compared OC and OC plus OS *MRI unit*: 1.5 T; receive-only shoulder surface coil *Sequences*: T1- and T2-weighted double-echo or fast spin-echo *Planes*: OC and OS NB: only proton-density and T2-weighted images obtained in the OC plane were reviewed	*Ref. test*: Arthroscopy/arthrotomy (21), arthrography (4), response to conservative treatment over 6 months (25). No further details reported *Test interval*: Not reported. At least 2 weeks between interpretation of MRI and US *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis provided (no reference given). Forced choice of positive or negative signs of tear using both signal intensity and morphologic criteria *Test interpreters*: 2 independent radiologists with 5 and 1 years MR experience

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Quinn, 1995[84] Radiology USA (1991–3) Retrosp Recruitment not reported Popl: 535 patients/ index: 535 shoulders/ ref. test: 100 shoulders *Mean age*: 47 (20–74) *Sex*: 54% male *Outcomes*: Any; FT; PT *Prevalence*: 31; 20; 11%	Symptomatic patients with clinically suspected subacromial impingement who underwent both index and reference test All patients had previously undergone conservative treatment with anti-inflammatory agents, restricted activity and physical therapy	*MRI unit*: 1.5 T; angled surface coil *Sequences*: Conventional fat-suppressed spin-echo with repetition time/echo time: 1. 2500/20; 2. 2000/20 (92 patients) or fat-suppressed fast spin-echo with repetition time/echo time: 3. 2000/80; 4. 2000/20 (8) *Planes*: 1. OC and AX; 2. OS; 3. OC and OS; 4. AX	*Ref. test*: Arthroscopy directed by clinical diagnosis: bursal surface only (50); bursal plus articular surface (50). Bursal if typical of impingement syndrome. Both if indeterminate history and physical findings. Authors acknowledge this may have overestimated sensitivity for partial tears but not for detection of complete tears *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis were provided (no reference given) *Test interpreters*: 4 radiologists with varying degrees of experience. Prospective interpretations used for study *Ref. test interpretation*: 9 orthopaedic surgeons with experience in shoulder arthroscopy who had access to the MR interpretations before arthroscopy. Written arthroscopy reports reviewed for study and in 18 patients arthroscopy videos were also reviewed (was attempted in all cases with positive MR or MR imaging interpretation error, but some videos were not available for review)
Reinus, 1995[86] Radiology US (1-year period) Cannot tell design Other[h] Popl: >200 patients/ index: >200 shoulders/ ref. test: 49[i] shoulders *Mean age*: 37 (28–65) *Sex*: [i]67% male *Outcomes*: FT; PT *Prevalence*: 20; 41%	Patients with shoulder pain referred specifically for MR imaging who underwent arthroscopy [h] State 'consecutive' yet 49/200 'selected'. Basis for selection presumably arthroscopy, but not explicitly stated	MRI: compared conventional and fat-suppressed imaging *MRI unit*: 1.5 T; shoulder coil *Sequences*: 1. T1-weighted and T2-weighted dual-echo conventional spin-echo sequences; 2. T2 weighted dual-echo spin-echo sequences with frequency selective presaturation of fat; 3. gradient echo *Planes*: 1. and 2. OC; 3. AX (not clear if latter were included in both sets or not)	*Ref. test*: Arthroscopy. Visualisation of the surfaces of the rotator cuff (joint, bursal or both) was at discretion of arthroscopist. Where an FT RCT was not seen on visualisation of one side of the RC, both sides were visualised *Test interval*: Mean 10 weeks (range 2 days to 66 weeks). 54% of MR exams were done within 8 weeks of arthroscopy, 22% had arthroscopy over 20 weeks after MRI *Further investigations*: None reported *Diagnostic criteria*: Criteria for diagnosis were agreed upon before initiation of the review and were provided (no reference given) *Test interpreters*: Experienced board-certified musculoskeletal radiologists ($n = 2$). Blinded to all clinical and surgical information *Ref. test interpretation*: Orthopaedic surgeons ($n = 6$). Three surgeons accounted for 84% of the cases (21, 13 and 7 cases). Radiologist reviewed MR again in conjunction with arthroscopy reported. Not clearly stated if this, or original AS diagnosis was used for results

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Robertson, 1995[85] Radiology/orthopaedics US (1989–91) Retrosp Recruitment not reported Popl: 97 patients/ index: 97[j] shoulders/ ref. test: 66[k] shoulders *Mean age*: [j]45 (17–72) *Sex*: 55% male *Outcomes*: Any; FT; PT *Prevalence*: 57; 31; 26%	Patients who had undergone surgery for RC disease plus asymptomatic volunteers and patients with instability Only those who underwent surgery within 6 months of MRI were included in the analysis – likely to be more severe cases [j] included 16 volunteers who did not get surgery [k] 7 with recurrent dislocation on basis of clin. exam. were excluded	*MRI unit*: 1.5 T; transmit–receive shoulder coil *Sequences*: 1. T1-weighted double-echo (48); 2. T1-weighted, fat-suppressed T2-weighted and T2 weighted (49) *Planes*: 1 and 2. Coronal and sagittal	*Ref. test*: Surgery. Open acromioplasty, direct visual evaluation and grading of the RC and RC repair when necessary had been undertaken *Test interval*: MR performed up to 6 months before surgery in 66 patients *Further investigations*: Not stated *Diagnostic criteria*: Criteria provided (no reference given) *Test interpreters*: Radiologists ($n = 4$): with 8 years, 3 years, 18 months and 4 weeks experience in musculoskeletal radiology. Images evaluated in random order independently and blinded to clinical information and reference test result *Ref. test interpretation*: Two authors
Sahin-Akyar, 1998[87] Radiology USA (1993–6) Retrosp Recruitment not reported Popl: 329 patients/ index: 329 shoulders/ ref. test: 39[l] shoulders *Mean age*: 54 (15–83) *Sex*: 44% male *Outcomes*: Any; FT; PT Prevalence: 62; 31; 31%	Patients referred for arthroscopy/surgery of the shoulder [l] AS performed in further 19, but MRI reports not available for review	MRI: compared standard gradient echo and fat-suppressed spin-echo *MRI unit*: 1.5 T *Sequences*: T1-weighted spin-echo imaging; fat-suppressed T2-weighted fast spin-echo; T2 weighted gradient-echo (all performed on all patients) *Planes*: OC	*Ref. test*: Arthroscopy or arthrotomy – both bursal and articular surfaces visualised *Test interval*: 2 days to 9 months, average 57 days *Further investigations*: Not reported *Diagnostic criteria*: Criteria for diagnosis based on reported criteria (incl. Zlatkin and several others) *Test interpreters*: Musculoskeletal radiologists with 2–10 years of experience ($n = 4$). Images reviewed independently and blinded to surgical findings twice (within 1 week) in two sets, each set containing T1-weighted images and either of the T2 images (which were randomised). Unaware of patient's name, age and sex *Ref. test interpretation*: Inferred retrospectively from written reports of 5 orthopaedic surgeons with extensive experience in shoulder arthroscopy and surgery. Surgeons had access to the official dictated report of the MR examination

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Swen, 1999[63] Rheumatology Netherlands (years not reported) Prosp2 Consecutive Popl: 21 patients/ index: 21 shoulders/ ref. test: 21 shoulders *Mean age*: 54 (SD 12) *Sex*: 57% male *Outcomes*: FT *Prevalence*: 62%	Consecutive patients awaiting surgery for clinically suspected RCT based on difficulty in abducting the arm, weakness and limitation of movement. Also had a lidocaine injection test into the acromion. All had unilateral non-inflammatory symptoms Neurological origins of shoulder weakness were excluded	Compared MRI and US (US details provided in Appendix 6) *MRI unit*: 1.0 T; shoulder coil *Sequences*: T1-weighted sequence plus standard T2 spin-echo sequence *Planes*: OC	*Ref. test*: Arthroscopy of the glenohumeral joint and the subacromial space. Subacromial decompression performed in all cases. FT RCTs repaired through a deltoid-splitting mini-incision *Test interval*: Within 3 weeks *Further investigations*: None reported *Diagnostic criteria*: Criteria provided (no reference given). *Test interpreters*: 'Experienced' radiologists (n = 2). Mean values of the 2 observers were used to estimate accuracy in paper, but results per reader also presented and extracted for this report *Ref. test interpretation*: 1 experienced shoulder surgeon
Torstensen, 1999[82] Sports medicine centre Canada (1993–95) Retrosp Recruitment not reported Popl: NR patients/ index: 57 shoulders/ ref. test: 57 shoulders *Mean age*: 41 (24–68) *Sex*: 58% male *Outcomes*: Any *Prevalence*: 42% male	Patients with shoulder pain who had undergone both MRI and arthroscopy. All failed conservative treatment MRI indicated if cause of pain unclear on clin. exam. (44) or if previous shoulder surgery without symptomatic improvement (13) Arthroscopy indicated if either history or clin. exam. consistent with impingement or instability, or if MRI significantly abnormal and inconclusive clinical assessment	*MRI unit*: 1.5 T (n = 34), 1.0 T (n = 16), 0.5 T (n = 7) (3 centres) *Sequences and planes*: A variety of different imaging sequences were said to have been used at each facility but were not reported	*Ref. test*: Arthroscopy. Included assessment of glenohumeral joint, labrum and biceps tendon and rotator cuff on both the subacromial and glenohumeral surfaces *Test interval*: Not described *Further investigations*: Not stated *Diagnostic criteria*: No details *Test interpreters*: 6 musculoskeletal radiologists (2 conducted 39/57 MRI reports). 35 patients had specialised shoulder arthroscopy reports available; these were reviewed along with MRI and operative findings *Ref. test interpretation*: 1 orthopaedic surgeon

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Traughber, 1992[114] Radiology/orthopaedics US (1989–90) Retrosp Recruitment not reported Popl: 40 patients/ index: 28 shoulders/ ref. test: 28 shoulders *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any; FT; PT; IS *Prevalence*: 50; 18; 32; 100%	Patients with clinical evidence of RC pathology who had undergone arthroscopic or combined open and arthroscopic surgery All had failed an extended course of conservative therapy	*MRI unit*: 1.5 or 0.5 T; surface coils *Sequences*: 1. Full shoulder evaluation (1.5 T): localiser series; spin-echo T1- and T2-weighted series; inversion recovery; and multiplanar gradient recall series. 2. Full shoulder exam. (0.5 T): localiser series, spin-echo T1- and T2-weighted series, inversion recovery series. 3. Screening RC exams: spin-echo T2-weighted; MPGR series and STIR series. *Planes*: 1. OC and AX; 2. OC; sagittal; AX; 3. OC; AX NB: there were additional changes in technique over time	*Ref. test*: Arthroscopic surgery or combined open and arthroscopic surgery. Where open surgery was elected, arthroscopy of the joint was performed first, with assessment of the articular surface of the cuff *Test interval*: Not reported *Further investigations*: None reported *Diagnostic criteria*: MR criteria for diagnosis of tears were described (no reference given). Examination protocols and reading criteria were established by two radiologists with experience in musculoskeletal imaging *Test interpreters*: Radiologist ($n = 4$)
Tuite, 1995,[99] 1994[115] Diagnostic radiology USA (1990–3) Retrosp Consecutive Popl: 100 patients/ index: 100 shoulders/ ref. test: 100[m] shoulders (59–24 cm FoV, 41–18 cm FoV). *Mean age*: 24 cm FoV: 43 18 cm FoV: 41 *Sex*: 71% male *Outcomes*: Any; FT; PT *Prevalence*: 56; 25; 31%	Consecutive patients who had both a shoulder MR and arthroscopy. Almost all had positive Neer test or other signs of impingement or RC disease [m] Subset of 87 patients reported in Tuite, 1994[115] (reports FT and PT tears separately) Tuite, 1995[99] tried to compare 2 different FoVs but did not use both in all patients. Have combined results to give bigger sample size	*MRI unit*: 1.5 T; single-loop dedicated shoulder coil *Sequences*: 1. T1-weighted; 2. gradient echo (GRE); 3. T2-weighted *Planes*: 1 and 2. OC; 3. OS	*Ref. test*: Arthroscopic evaluation using a standardised technique with the surgeon having knowledge of the MRI findings *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Articular and bursal surfaces graded using an MR equivalent of the arthroscopic grading system. Tuite, 1994[115] references previously published criteria for GRE imaging (Tuite, 1994;[115] Holt, 1990) *Test interpreters*: Two of the authors reviewed jointly the images at the end of the study period and a consensus reading obtained without knowledge of the history or arthroscopic results *Ref. test interpretation*: One orthopedist subspeciality trained in shoulder surgery, with a practice specialising in shoulder disease. Not independent of MRI findings

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Tuite, 2001[93] Radiology USA (1998–9) Retrosp Consecutive Popl: 344 patients/ index: 344 shoulders/ ref. test: 75[n] shoulders *Mean age*: 40 (16–64) *Sex*: [n]65% male *Outcomes*: Any *Prevalence*: 65%	Patients referred for shoulder MR scan who underwent arthroscopy and subacromial bursoscopy	MRI: compared standard OS and angled OS *MRI unit*: 1.5 T; phased-array shoulder coil *Sequences*: Fat-suppressed fast spin-echo T2-weighted *Planes*: 1. OC/OS images; 2. OC/AS NB: AS set had smaller section thickness which may have improved accuracy	*Ref. test*: Arthroscopy and subacromial bursoscopy *Test interval*: Average time 95 days (range 6–319). Plus 4 weeks between reading index tests 1 and 2 *Further investigations*: At same time had subacromial bursography *Diagnostic criteria*: Criteria for diagnosis were provided (Reinus, 1995[86]). Forced choice: torn (partial or full thickness) or not *Test interpreters*: Musculoskeletal radiologist and musculoskeletal radiology fellow with 4 months' experience. Masked to history and arthroscopic results *Ref. test interpretation*: Single experienced shoulder surgeon
Wang, 1994[96] Radiology and orthopaedics Taiwan (1990–2) Retrosp Consecutive Popl: 40 patients/ index: 40 shoulders/ ref. test: 40 shoulders *Mean age*: 49 (17–75) *Sex*: 67% male *Outcomes*: Any; FT; PT *Prevalence*: 55; 47; 8%	Patients with shoulder pain who underwent surgery	*MRI unit*: 1.5 T; planar surface coil *Sequences*: 1. TR/TE = 1800–2000/20,80; 2. T1-weighted sequence and a multiplanar gradient recalled pulse sequence (axial) *Planes*: 1. OC and OS; 2. AX	*Ref. test*: Arthroscopic surgery (for 12 patients with clinical indication of instability); open surgery was performed on the remainder *Test interval*: Within 1 month *Further investigations*: None reported *Diagnostic criteria*: Criteria for diagnosis provided (no reference given) *Test interpreters*: Radiologists (n = 2). Blinded to patient name and surgical results

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Wnorowski, 1997[37] Orthopaedic centre USA (1990–4) Retrosp Recruitment not reported Popl: NR° patients/index: 38 shoulders (39 shoulders)/ ref. test: 39 shoulders *Mean age*: Median 30 (20–75) *Sex*: Not reported *Outcomes*: Any, FT, PT, *Prevalence*: 37; 23; 13%	Patients seen for a shoulder problem (and who underwent MRI) The majority of cases were relatively difficult, referrals from other surgeons, or diagnostic dilemmas (the primary diagnosis was unclear after clin. exam.) ° 20% of those undergoing arthroscopy during time period underwent MRI and were included	Compared MRI and clin. exam. (clin. exam. details are reported in Appendix 5) *MRI unit*: 1.5 T; specialist shoulder coil *Sequences*: T1-weighted, T2-weighted and GRE images *Planes*: OS, transverse and OC	*Ref. test*: Arthroscopy and subacromial bursal evaluation of the cuff *Test interval*: Mean interval 2 months, median 12 weeks. Two patients had a delay of 2 years *Further investigations*: In those for whom MRI indicated a tear but arthroscopy no tear or partial tear, an 'arthroscopic arthrogram' was performed. Appropriate therapeutic interventions were then started. All cases had clin. exam., either at time of MRI (7) or before MRI (31) *Diagnostic criteria*: No criteria provided *Test interpreters*: MRI undertaken at 1 of 3 centres in the area, depending on geographic considerations, physician preference and/or insurance requirements. Radiologists in the CCS (n = 12), one experienced musculoskeletal radiologist (ESS), and one study author. 4 CCS radiologists read first 31 images, then 1 each for last 8 patients. Original radiologists reports were used and classified retrospectively
Wolf, 2001[38] Medical Centre USA (1999–2000) Cannot tell design Consecutive Popl: 109 patients/index: 109 shoulders (71 got MRI)/ref. test: 109 shoulders *Mean age*: 51.2 (29–86) *Sex*: 61% male *Outcomes*: FT *Prevalence*: 46%	Patients undergoing arthroscopy for diagnoses relating to shoulder pain and weakness	Evaluated clin. exam and MRI (clin. exam. details are given in Appendix 5) *MRI unit*: No details *Sequences*: No details	*Ref. test*: Arthroscopy *Test interval*: Not reported *Further investigations*: Not stated *Diagnostic criteria*: No details given *Test interpreters*: Not reported. Retrospective chart review used to document clinical, surgical and MRI findings. MRI evaluated independently of clinical examination

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Yagci, 2001[97] Radiology/orthopaedics Turkey (1997–8) Prosp2 Recruitment not reported Popl: NR patients/index: 24 shoulders/ref. test: 24 shoulders *Mean age*: 52 (16–73) *Sex*: 29% male *Outcomes*: Any; FT; PT *Prevalence*: 62; 42; 21%	Clinically suspected labral or RC abnormalities. All referred for surgery (therapeutic)	Evaluated MRI and MRA (MRA details are given in Appendix 12) *MRI unit*: 1.0 T; shoulder surface coil *Sequences*: 1. T1-weighted; 2. T2-weighted images; 3. T2-weighted GRE fast imaging with steady-state precession images *Planes*: 1. AX and PC; 2. PC; 3. PS	*Ref. test*: Arthroscopy (n = 1) or open surgery (n = 24) *Test interval*: Between MRI and MRA, mean 11 days (range 0–27 days). Between MRA and surgery 7 days (range 1–23 days) *Further investigations*: Not reported *Diagnostic criteria*: Standard criteria established in recent literature were used (no references given) *Test interpreters*: Experienced musculoskeletal radiologists (n = 2)
Zlatkin, 1989[16] Radiology/orthopaedics USA (1987–8) Cannot tell design Consecutive 40/40/32[p] *Mean age*: Not reported *Sex*: Not reported *Outcomes*: Any *Prevalence*: 69%	Surgical patients with a clinical diagnosis of RC tendinopathy. Surgery indicated on basis of symptoms or disability All had failed standardised conservative treatment and had positive impingement test with at least 75% subjective short-term improvement in rest pain and in pain associated with the impingement signs [p] 8 volunteers included. Note later states 14/24 subjects without RCT were asymptomatic volunteers	*MRI unit*: 1.5 T; two general-purpose receive-only surface coils *Sequences*: 1. Short repetition/short echo time (T1-weighted); 2. long TR/TE (T2-weighted) *Planes*: 1. AX, OS and OC; 2. OC. NB: true coronal and sagittal only in 8 patients	*Ref. test*: Surgery: open acromioplasty and subacromial decompression in all cases. 5 patients with intact RCs at surgery also underwent inspection of the cuff with arthroscopy, as did 3 patients with partial tears. *Test interval*: Not reported *Further investigations*: 24 patients underwent arthrography *Diagnostic criteria*: Criteria for diagnosis based in part on those developed by one of the authors and his co-workers in a previous study (Zlatkin, 1988).[147] Final consensus diagnosis often made on basis of the readers' overall impression rather than strict adherence to the above criteria. To try to increase objectivity, a scoring system was retrospectively developed based on the appearance of the tendon and the status of the subacromial-subdeltoid fat plane *Test interpreters*: Radiologists with experience of musculoskeletal MR imaging (n = 3). Blinded to name, clinical history, surgical reports and other studies. MR reports for asymptomatic volunteers randomly included *Ref. test interpretation*: Orthopaedic surgeons. Arthrograms stated to have been reviewed by 2/3 radiologists

Appendix 10

MRI: sensitivity and specificity results

Study	Comparison[a]	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Balich, 1997[83]	R1[(L)]	0.69 (0.58 to 0.79)	0.87 (0.80 to 0.91)	0.84 (0.71 to 0.92)	0.96 (0.92 to 0.98)	0.35 (0.19 to 0.54)	0.97 (0.91 to 0.99)
	R2	0.70 (0.59 to 0.80)	0.81 (0.75 to 0.87)	0.87 (0.74 to 0.94)	0.94 (0.89 to 0.96)	0.42 (0.26 to 0.61)	0.89 (0.82 to 0.93)
	R3[b]	0.69 (0.58 to 0.79)	0.95 (0.90 to 0.97)	0.89 (0.77 to 0.95)	0.98 (0.94 to 0.99)	0.42 (0.26 to 0.61)	0.85 (0.78 to 0.91)
	R4	0.76 (0.65 to 0.84)	0.94 (0.89 to 0.97)	0.96 (0.85 to 0.99)	0.97 (0.94 to 0.99)	0.42 (0.26 to 0.61)	0.97 (0.91 to 0.99)
	R5[(M)]	0.76 (0.65 to 0.84)	0.94 (0.89 to 0.97)	0.96 (0.85 to 0.99)	0.97 (0.93 to 0.98)	0.42 (0.26 to 0.61)	0.97 (0.93 to 0.99)
Birtane, 2001[88]		0.99[I] (0.94 to 1.00)	0.37[I] (0.23 to 0.53)	–	–	–	–
Blanchard, 1999[98]		–	–	0.82 (0.52 to 0.95)	0.78 (0.59 to 0.89)	–	–
Burk, 1989[71]		1.00 (0.85 to 1.00)	0.88 (0.64 to 0.97)	–	–	–	–
Evancho, 1988[89]		0.69 (0.42 to 0.87)	0.94 (0.74 to 0.99)	0.80 (0.49 to 0.94)	0.94 (0.74 to 0.99)	0.33 (0.6 to 0.79)	0.94 (0.74 to 0.99)
Hodler, 1991[59]		–	–	0.67 (0.42 to 0.85)	0.89 (0.56 to 0.98)	–	–
Hodler, 1992[90]		0.41 (0.22 to 0.64)	0.79 (0.57 to 0.91)	1.00 (0.51 to 1.00)	0.88 (0.72 to 0.95)	0.8 (0.1 to 0.33)	0.91 (0.73 to 0.98)
Iannotti, 1991[17]		0.84[I] (0.67 to 0.93)	0.96[I] (0.88 to 0.99)	1.00 (0.90 to 1.00)	0.95 (0.85 to 0.98)	–	–
Jaovisidha, 1999[91]		1.00 (0.34 to 100)	1.00 (0.61 to 1.00)	–	–	–	–
Kieft, 1988[113]		100 (0.44 to1.00)	0.71 (0.36 to 0.92)	–	–	–	–
Kneeland, 1987[95]		0.91 (0.72 to 0.97)	0.75 (0.30 to 0.95)	–	–	–	–
Martin-Hervas, 2001[66]		0.91 (0.77 to 0.97)	0.74 (0.55 to 0.87)	0.81 (0.62 to 0.91)	0.97 (0.85 to 0.99)	0.50 (0.22 to 0.78)	0.75 (0.62 to 0.85)
Morrison, 1990[94]		1.00 (0.93 to 1.00)	0.96 (0.85 to 0.99)	1.00 (0.93 to 1.00)	0.88 (0.76 to 0.94)	–	–
Nelson, 1991[64]		0.88[I] (0.53 to 0.98)	0.77[I] (0.50 to 0.92)	0.86 (0.49 to 0.97)	0.93 (0.69 to 0.99)	0.67 (0.39 to 0.86)	0.89 (0.56 to 0.98)
Patten, 1994[92]	OC	0.85 (0.64 to 0.95)	0.80 (0.63 to 0.90)	–	–	–	–
	OC, OS[b]	0.95 (0.76 to 0.99)	0.93 (0.79 to 0.98)	–	–	–	–
Quinn, 1995[84]		0.84 (0.67 to 0.93)	0.97 (0.90 to 0.99)	0.85 (0.64 to 0.95)	0.99 (0.93 to 1.00)	0.82 (0.52 to 0.95)	0.99 (0.94 to 1.00)

continued

Study	Comparison[a]	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Reinus, 1995[86]	C: R1	–	–	0.80 (0.49 to 0.94)	0.87 (0.73 to 0.94)	0.10 (0.3 to 0.30)	0.93 (0.78 to 0.98)
	C: R2[b]	–	–	0.80 (0.49 to 0.94)	0.92 (0.80 to 0.97)	0.20 (0.8 to 0.42)	0.83 (0.65 to 0.92)
	F: R1[b]	–	–	1.00 (0.72 to 1.00)	0.77 (0.62 to 0.87)	0.45 (0.26 to 0.66)	0.93 (0.78 to 0.98)
	F: R2	–	–	1.00 (0.72 to 1.00)	0.87 (0.73 to 0.94)	0.25 (0.11 to 0.47)	0.97 (0.83 to 0.99)
Robertson, 1995[85]	R1[(M)b]	–	–	0.96 (0.80 to 0.99)	0.98 (0.91 to 1.00)	0.57 (0.37 to 0.76)	0.85 (0.74 to 0.92)
	R2	–	–	1.00 (0.87 to 1.00)	0.91 (0.81 to 0.96)	0.38 (0.21 to 0.59)	0.93 (0.84 to 0.97)
	R3	–	–	0.85 (0.66 to 0.94)	0.89 (0.79 to 0.95)	0.19 (0.8 to 0.40)	0.92 (0.82 to 0.96)
	R4[(L)]	–	–	0.81 (0.62 to 0.91)	0.96 (0.88 to 0.99)	0.24 (0.11 to 0.45)	0.92 (0.82 to 0.96)
Sahin–Akyar, 1998[87]	C: R1[b]	0.71 (0.51 to 0.85)	0.80 (0.55 to 0.93)	0.83 (0.55 to 0.95)	0.96 (0.82 to 0.99)	0.42 (0.19 to 0.68)	0.85 (0.68 to 0.94)
	C: R2	0.88 (60.9 to 0.96)	0.60 (0.36 to 0.80)	1.00 (0.76 to 1.00)	0.85 (0.68 to 0.94)	0.50 (0.25 to 0.75)	0.81 (0.63 to 0.92)
	C: R3	0. 75 (0.55 to 0.88)	0.87 (0.62 to 0.96)	1.00 (0.76 to 1.00)	0.81 (0.63 to 0.92)	0.25 (0.9 to 0.53)	1.00 (0.88 to 1.00)
	C: R4	0.67 (0.47 to 0.82)	0.60 (0.36 to 0.80)	0.83 (0.55 to 0.95)	0.78 (0.59 to 0.89)	0.17 (0.5 to 0.45)	0.85 (0.68 to 0.94)
	F: R1[b]	0.96 (0.80 to 0.99)	0.60 (0.36 to 0.80)	0.83 (0.55 to 0.95)	0.93 (0.77 to 0.98)	0.75 (0.47 to 0.91)	0.70 (0.52 to 0.84)
	F: R2	0.92 (0.74 to 0.98)	0.33 (0.15 to 0.58)	0.92 (0.65 to 0.99)	0.93 (0.77 to 0.98)	0.67 (0.39 to 0.86)	0.59 (0.41 to 0.75)
	F: R3	0.79 (0.60 to 0.91)	0.87 (0.62 to 0.96)	1.00 (0.76 to 1.00)	0.81 (0.63 to 0.92)	0.25 (0.9 to 0.53)	0.96 (0.82 to 0.99)
	F: R4	0.83 (0.64 to 0.93)	0.87 (0.62 to 0.96)	0.83 (0.55 to 0.95)	0.89 (0.72 to 0.96)	0.50 (0.25 to 0.75)	0.89 (0.72 to 0.96)
Swen, 1999[63]	R1[b]	–	–	0.77 (0.50 to 0.92)	0.88 (0.53 to 0.98)	–	–
	R2	–	–	0.85 (0.58 to 0.96)	0.88 (0.53 to 0.98)	–	–
Torstensen, 1999[82]		0.96 (0.80 to 0.99)	0.48 (0.33 to 0.65)	–	–	–	–
Traughber, 1992[114]		0.71 (0.45 to 0.88)	0.93 (0.69 to 0.99)	1.00 (0.57 to 1.00)	1.00 (0.86 to 1.00)	0.56 (0.27 to 0.81)	0.95 (0.75 to 0.99)
Tuite, 1995[99]		0.93 (0.83 to 0.97)	0.75 (0.61 to 0.85)	0.91 (0.72 to 0.97)	0.95 (0.87 to 0.98)	0.74 (0.55 to 0.87)	0.87 (0.76 to 0.93)
Tuite, 2001[93]	OS: R1[(L)]	0.84 (0.71 to 0.91)	0.54 (0.35 to 0.71)	–	–	–	–
	OS: R2[(M)b]	0.73 (0.60 to 0.84)	0.81 (0.62 to 0.91)	–	–	–	–
	AS: R1	0.86 (0.73 to 0.93)	0.92 (0.76 to 0.98)	–	–	–	–
	AS: R2	0.88 (0.76 to 0.94)	0.54 (0.35 to 0.71)	–	–	–	–

continued

Study	Comparison[a]	Any tear		Full tear		Partial tear	
		Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)	Se (95% CI)	Sp (95% CI)
Wang, 1994[96]		0.91 (0.72 to 0.97)	0.89 (0.67 to 0.97)	–	–	–	–
Wnorowski, 1997[37]	CCS$^{(L)}$	0.86 (0.60 to 0.96)	0.52 (0.33 to 0.70)	0.56 (0.27 to 0.81)	0.73 (0.56 to 0.86)	0.33 (0.6 to 0.79)	1.00 (0.91 to 1.00)
	ESS$^{(M)}$[b]	0.71 (0.45 to 0.88)	0.71 (0.51 to 0.85)	0.78 (0.45 to 0.94)	0.83 (0.65 to 0.92)	0 (0 to 0.43)	0.68 (0.51 to 0.81)
Wolf, 2001[38]		–	–	0.91 (0.76 to 0.97)	0.89 (0.76 to 0.96)	0.20 (0.4 to 0.62)	0.88 (0.73 to 0.95)
Yagci, 2001[97]	R1[b]	0.53 (0.30 to 0.75)	0.56 (0.27 to 0.81)	0.80 (0.49 to 0.94)	0.71 (0.45 to 0.88)	0 (0 to 0.43)	1.00 (0.83 to 1.00)
	R2	0.60 (0.36 to 0.80)	0.33 (0.12 to 0.65)	0.90 (0.60 to 0.98)	0.64 (0.39 to 0.84)	0 (0 to 0.43)	0.95 (0.75 to 0.99)
Zlatkin, 1989[16]		0.91 (0.72 to 0.97)	0.70 (0.40 to 0.89)	–	–	–	–

Se, sensitivity; Sp, specificity.
[a] R, reader (L, least experienced; M, most experienced) ; C, conventional MR; F, fat-suppressed MR; OC, oblique coronal view; OS, oblique sagittal view; AS, angled sagittal view; CCS, clinical community setting radiologists; ESS, expert radiologists.
[b] Indicates set of results used for meta-analysis.
[l] Data for 'impingement syndrome' not for 'any tear'.

Appendix 11

MRI: positive and negative likelihood ratios

Study	Comparison[a]	Any tear		Full tear		Partial tear	
		LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)
Balich, 1997[83]	R1[(L)]	5.2 (3.4 to 8.1)	0.4 (0.3 to 0.5)	21.4 (10.2 to 44.6)	0.2 (0.1 to 0.3)	10.0 (3.3 to 29.8)	0.7 (0.5 to 0.9)
	R2	3.8 (2.6 to 5.5)	0.4 (0.3 to 0.5)	13.9 (7.8 to 25.0)	0.1 (0.1 to 0.3)	3.7 (1.9 to 7.4)	0.7 (0.5 to 0.9)
	R3[b]	13.0 (6.5 to 26.0)	0.3 (0.2 to 0.5)	39.3 (14.8 to 104.2)	0.1 (0.0 to 0.3)	2.9 (1.5 to 5.4)	0.7 (0.5 to 0.9)
	R4	12.8 (6.7 to 24.4)	0.3 (0.2 to 0.4)	33.8 (14.2 to 80.4)	0.0 (0.0 to 0.2)	12.2 (4.2 to 35.2)	0.6 (0.4 to 0.8)
	R5[(M)]	12.8 (6.7 to 24.4)	0.3 (0.2 to 0.4)	28.3 (12.9 to 62.4)	0.0 (0.0 to 0.2)	16.2 (4.9 to 54.0)	0.6 (0.4 to 0.8)
Birtane, 2001[88]		1.6 (1.2 to 2.0)[l]	0.0 (0.0 to 0.2)[l]	–	–	–	–
Blanchard, 1999[98]		–	–	3.7 (1.7 to 7.9)	0.2 (0.1 to 0.8)	–	–
Burk, 1989[71]		8.0 (2.2 to 29.2)	0.0 (0.0 to 0.4)	–	–	–	–
Evancho, 1988[89]		12.5 (1.8 to 86.6)	0.3 (0.1 to 0.7)	14.4 (2.1 to 99.2)	0.2 (0.1 to 0.7)	6.0 (0.5 to 72.2)	0.7 (0.3 to 1.6)
Hodler, 1991[59]		–	–	6.0 (0.9 to 39.4)	0.4 (0.2 to 0.8)	–	–
Hodler, 1992[90]		2.0 (0.7 to 5.5)	0.7 (0.5 to 1.2)	8.0 (3.2 to 20.0)	0.1 (0.0 to 0.6)	0.9 (0.1 to 8.8)	1.0 (0.8 to 1.2)
Iannotti, 1991[17]		23.9 (6.1 to 94.1)[l]	0.2 (0.1 to 0.4)[l]	18.3 (6.1 to 55.1)	0.0 (0.0 to 0.2)	–	–
Jaovisidha, 1999[91]		11.7 (0.8 to176.8)	0.2 (0.0 to 2.3)	–	–	–	–
Kieft, 1988[113]		3.5 (1.1 to 11.3)	0.2 (0.0 to 2.5)	–	–	–	–
Kneeland, 1987[95]		3.6 (0.7 to 20.0)	0.1 (0.0 to 0.5)	–	–	–	–
Martin–Hervas, 2001[66]		3.5 (1.8 to 6.7)	0.1 (0.0 to 0.4)	28.3 (4.1 to 196.9)	0.2 (0.1 to 0.4)	2.0 (0.9 to 4.7)	0.7 (0.3 to 1.3)
Morrison, 1990[94]		22.5 (5.8 to 87.2)	0.0 (0.0 to 0.1)	8.2 (3.9 to 17.3)	0.0 (0.0 to 0.2)	–	–
Nelson, 1991[64]		3.8 (1.4 to 10.6)[l]	0.2 (0.0 to 1.0)[l]	12.0 (1.8 to 81.3)	0.2 (0.0 to 0.9)	6.0 (0.9 to 39.7)	0.4 (0.2 to 0.9)
Patten, 1994[92]	OC	4.3 (2.0 to 8.9)	0.2 (0.1 to 0.5)	–	–	–	–
	OC, OS[b]	14.3 (3.7 to 54.6)	0.1 (0.0 to 0.4)	–	–	–	–
Quinn, 1995[84]		28.9 (7.3 to 114.4)	0.2 (0.1 to 0.4)	68.0 (9.6 to 481.0)	0.2 (0.1 to 0.4)	72.8 (10.2 to 521.5)	0.2 (0.1 to 0.6)

continued

Study	Comparison[a]	Any tear		Full tear		Partial tear	
		LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)
Reinus, 1995[86]	C: R1	–	–	6.2 (2.6 to 15.0)	0.2 (0.1 to 0.8)	1.5 (0.2 to 9.5)	1.0 (0.8 to 1.2)
	C: R2[b]	–	–	10.4 (3.4 to 32.2)	0.2 (0.1 to 0.8)	1.2 (0.4 to 3.8)	1.0 (0.7 to 1.3)
	F: R1[b]	–	–	4.3 (2.4 to 7.7)	0.1 (0.0 to 0.9)	6.5 (1.6 to 27.1)	0.6 (0.4 to 0.9)
	F: R2	–	–	7.8 (3.4 to 17.7)	0.0 (0.0 to 0.8)	7.3 (0.9 to 57.5)	0.8 (0.6 to 1.0)
Robertson, 1995[85]	R1[(M)b]	–	–	53.8 (7.7 to 375.6)	0.0 (0.0 to 0.3)	3.8 (1.9 to 7.7)	0.3 (0.8 to 7.6)
	R2	–	–	11.2 (4.9 to 25.9)	0.0 (0.0 to 0.3)	5.8 (1.9 to 17.3)	0.7 (0.5 to 0.9)
	R3	–	–	7.9 (3.6 to 17.1)	0.2 (0.1 to 0.4)	2.3 (0.7 to 7.9)	0.9 (0.7 to 1.1)
	R4[(L)]	–	–	22.6 (5.7 to 89.3)	0.2 (0.1 to 0.4)	2.9 (0.9 to 9.0)	0.8 (0.6 to 1.1)
Sahin–Akyar, 1998[87]	C: R1[b]	3.5 (1.2 to 10.1)	0.4 (0.2 to 0.7)	22.5 (3.2 to 156.6)	0.2 (0.0 to 0.6)	2.8 (0.9 to 8.7)	0.7 (0.4 to 1.1)
	C: R2	2.2 (1.2 to 4.1)	0.2 (0.1 to 0.6)	6.8 (2.7 to 16.7)	0.0 (0.0 to 0.7)	2.7 (1.0 to 7.1)	0.6 (0.3 to 1.1)
	C: R3	5.6 (1.5 to 20.9)	0.3 (0.1 to 0.6)	5.4 (2.4 to 11.9)	0.0 (0.0 to 0.7)	6.8 (0.8 to 58.4)	0.8 (0.5 to 1.0)
	C: R4	1.7 (0.8 to 3.3)	0.6 (0.3 to 1.1)	3.8 (1.8 to 7.9)	0.2 (0.1 to 0.8)	1.1 (0.2 to 5.3)	1.0 (0.7 to 1.3)
	F: R1[b]	2.4 (1.3 to 4.5)	0.1 (0.0 to 0.5)	11.3 (2.9 to 43.7)	0.2 (0.1 to 0.6)	2.5 (1.3 to 4.9)	0.4 (0.1 to 1.0)
	F: R2	1.4 (0.9 to 2.0)	0.3 (0.1 to 1.1)	12.4 (3.2 to 47.5)	0.1 (0.0 to 0.6)	1.6 (0.9 to 3.0)	0.6 (0.2 to 1.3)
	F: R3	5.9 (1.6 to 21.9)	0.2 (0.1 to 0.5)	5.4 (2.4 to 11.9)	0.0 (0.0 to 0.7)	6.7 (0.8 to 58.5)	0.8 (0.6 to 1.1)
	F: R4	6.3 (1.7 to 23.0)	0.2 (0.1 to 0.5)	7.5 (2.5 to 22.5)	0.2 (0.1 to 0.7)	4.5 (1.3 to 15.1)	0.6 (0.3 to 1.0)
Swen, 1999[63]	R1[b]	–	–	6.2 (1.0 to 39.4)	0.3 (0.1 to 0.7)	–	–
	R2			6.8 (1.1 to 43.0)	0.2 (0.0 to 0.6)		
Torstensen, 1999[82]		1.9 (1.3 to 2.6)	0.1 (0.0 to 0.6)	–	–	–	–
Traughber, 1992[114]		10.0 (1.5 to 68.0)	0.3 (0.1 to 0.7)	44 (2.8 to 690.6)	0.1 (0.0 to 1.2)	10.6 (1.4 to 77.6)	0.5 (0.2 to 1.0)
Tuite, 1995[99]		3.7 (2.2 to 6.2)	0.1 (0.0 to 0.2)	19.7 (6.5 to 60.0)	0.1 (0.0 to 0.4)	5.6 (2.8 to 11.0)	0.3 (0.2 to 0.6)
Tuite to 2001[93]	OS: R1[(L)]	1.8 (1.2 to 2.8)	0.3 (0.1 to 0.6)				
	OS: R2[(M)b]	3.8 (1.7 to 8.5)	0.3 (0.2 to 0.5)	–	–	–	–
	AS: R1	1.9 (1.2 to 2.9)	0.2 (0.1 to 0.5)				
	AS: R2	11.1 (2.9 to 42.4)	0.2 (0.1 to 0.3)				

continued

Study	Comparison[a]	Any tear		Full tear		Partial tear	
		LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)	LR+ (95% CI)	LR– (95% CI)
Wang, 1994[96]		8.2 (2.2 to 30.4)	0.1 (0.0 to 0.4)	–	–	–	–
Wnorowski, 1997[37]	CCS$^{(L)}$	1.8 (1.1 to 2.8)	0.3 (0.1 to 1.0)	2.1 (0.9 to 4.8)	0.6 (0.3 to 1.3)	28.5 (1.4 to 590.0)	0.7 (0.3 to 1.5)
	ESS(M)[b]	2.4 (1.2 to 5.0)	0.4 (0.2 to 1.0)	4.5 (1.9 to 10.8)	0.3 (0.1 to 0.9)	0.3 (0.0 to 3.8)	1.5 (1.2 to 1.9)
Wolf, 2001[38]		–	–	8.6 (3.4 to 22.0)	0.1 (0.0 to 0.3)	1.7 (0.2 to 11.9)	0.9 (0.6 to 1.4)
Yagci, 2001[97]	R1[b]	1.2 (0.5 to 2.9)	0.8 (0.4 to 1.9)	2.8 (1.2 to 6.8)	0.3 (0.1 to 1.0)	–[c]	–[c]
	R2	0.9 (0.5 to 1.7)	1.2 (0.4 to 3.7)	2.5 (1.2 to 5.2)	0.2 (0.0 to 1.0)	1.1 (0.1 to 23.9)	1.1 (0.9 to 1.2)
Zlatkin, 1989[16]		3.0 (1.2 to 7.9)	0.1 (0.0 to 0.5)	–	–	–	–

LR+, positive likelihood ratio; LR–, negative likelihood ratio.
[a] R, reader (^{L}, least experienced; ^{M}, most-experienced); C, conventional MR; F, fat-suppressed MR; OC, oblique coronal view; OS, oblique sagittal view; AS, angled sagittal view; CCS, clinical community radiologists; ESS, expert radiologists.
[b] Indicates set of results used for meta-analysis.
[c] Not possible to estimate since sensitivity was 0.
[l] Data for 'impingement syndrome' not for 'any tear'.

Appendix 12

MRA: detailed study methods

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Binkert, 2001[103] Radiology/orthopaedics Switzerland (1999–2000) Prosp3 Consecutive Popl: 156 patients/index: 156 shoulders (52 in each solution)/ ref. test: 88 shoulders(30 in 2 mmol, 26 in 4 mmol, 32 in Ringer) *Mean age*: 51 (14–83) *Sex*: 59% male *Outcomes*: Any *Prevalence*: 35%	Patients referred for MRA of glenohumeral joint Prior shoulder surgery or arthroscopy excluded	Evaluated MRA with three different contrast solutions *MRI unit*: 1.0 T; dedicated receive-only shoulder coil *Sequences*: 1. T1-weighted; 2. dual spin-echo and T1 fat-suppressed images. *Planes*: 1. transverse and PS images; 2. AC *Contrast*: 2 mmol gadoteridol contrast agent; 4 mmol gadoteridol contrast agent; Ringer solution contrast agent	*Ref. test*: Arthroscopy (62)/open surgery (26). No further details *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Established criteria from literature used to differentiate the various abnormalities (references provided) *Test interpreters*: Two independent staff radiologists specialising in musculoskeletal radiology. Blinded to clinical information and to gadoteridol concentration but could not be blinded to Ringer solution owing to different sequences used *Ref. test interpretation*: Not reported
Funke, 1996[101] Radiology Germany (years not reported) Prosp2 Recruitment not reported Popl: NR patients/ index: 25 shoulders/ ref. test: 25 shoulders *Mean age*: 43 (19–64) *Sex*: 64% male *Outcomes*: Any; FT; PT *Prevalence*: 68; 56; 12%	Patients referred for evaluation of RC	Evaluated MRA using either standard spin-echo MRI or spin-echo fat-suppressed MRI *MRI unit*: 1.5 T; dedicated shoulder coil *Sequences*: Standard spin-echo; spin-echo fat-suppressed *Planes*: OC, and OS; OC *Contrast*: 1 ml gadopentetate dimeglumine and 10 ml iotrolan	*Ref. test*: Arthrography (single-contrast). Details provided *Test interval*: Arthrography given first, then MRA followed *Further investigations*: Three patients underwent arthroscopy, 4 had open surgery. In patients who did not undergo surgery, the results of the conventional arthrography in combination with both MR techniques were used to establish a final diagnosis (results not extracted) *Diagnostic criteria*: Criteria reported *Test interpreters*: Panel of 2 experienced radiologists. Standard MR interpreted blinded to clinical information, fat-suppressed MR interpreted alongside standard images. Diagnosis established without knowledge of arthrography *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Hodler, 1992[90] Radiology/orthopaedics USA (years not reported) Cannot tell design Recruitment not reported Popl: 150 patients/ index: 36• shoulders/ ref. test: 36 shoulders *Mean age*: 42.5 (17–69) *Sex*: 67% male *Outcomes*: Any; FT; PT *Prevalence*: 47; 11; 36%	Patients who had MRI and whose arthroscopy reports were available Selected sub-sample of those undergoing MRI for MRA – no reasons for selection were given	Evaluated MRI (see Appendix 9) and MRA *MRI unit*: 1.5 T magnet *Sequences*: proton-density, T2-weighted; T1-weighted used for MRA *Planes*: OC; OC and OS used for MRA *Contrast*: 1 ml of gadopentetate dimeglumine, 469 mg ml^{-1} in 250 ml saline	*Ref. test*: Arthroscopy: details provided. Examined glenohumeral and subacromial bursa ($n = 32$). 3 young patients with shoulder instability and 1 with adhesion capsulitis did not undergo bursal examination. *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Criteria provided (no reference given) *Test interpreters*: Experienced osteoradiologists ($n = 3$) *Ref. test interpretation*: Experienced surgeons
Loew, 2000[102] Radiology Germany (years not reported) Prosp3 Recruitment not reported Popl: 38 shoulders/index: [a]38 shoulders/ref. test: 27 shoulders *Mean age*: 48 (21–80) *Sex*: [a]66% male *Outcomes*: FT; *Prevalence*: 44%	Patients with suspected chronic instability or RC abnormalities	Evaluated MRA using either low-field MRI or fat-suppressed high-field MRI *MRI unit*: 0.2 or 1.5 T; flexible surface coil *Sequences*: Not reported for low field; high field used fat suppression in T1-weighted images. Discussion reports use of different sequences at each MR unit *Planes*: Not reported *Contrast*: Gadopentetate dimeglumine MRI started within 15 minutes of arthrography and was performed in randomised order: low then high field	*Ref. test*: Arthroscopy and surgery in all patients. Arthroscopic evaluation of the labrum and RC performed before surgery *Test interval*: Not reported *Further investigations*: Not reported *Diagnostic criteria*: Reported for high field *Test interpreters*: Radiologists: one musculoskeletal radiologist with 5 years' post-fellowship experience and one fifth-year resident with 2 years' MRI experience. Blinded to patient data, clinical history and surgical results. Both readers had same results, so only one set extracted *Ref. test interpretation*: Not reported

continued

Study details	Eligibility criteria	Index test(s) details	Reference test and test interpretation
Pfirrmann, 1999[100] Radiology/orthopaedics Switzerland (years not reported) Retrosp Consecutive Popl: 50 patients/ index: 50 shoulders/ ref. test: 50 shoulders *Mean age*: 50.6 (21–76) *Sex*: 70% male *Outcomes*: Any *Prevalence*: 42%	Patients undergoing MRA at authors' institution and surgery from a single shoulder surgeon Excluded if previous surgery	Evaluated MRA using three different views *MRI unit*: 1.0 T; receive-only shoulder coil *Sequences*: T1-weighted turbo spin-echo images *Planes*: PS, transverse and AC; transverse view only; PS view only. *Contrast*: 10 ml diluted gadopentetate dimeglumine (concentration 4 mmol l^{-1})	*Ref. test*: Arthroscopy (21)/open surgery (29). No details provided *Test interval*: Max. 3 months for inclusion *Further investigations*: Not described *Diagnostic criteria*: Imaging criteria provided *Test interpreters*: Radiologists who were specialised in musculoskeletal radiology independently and unaware of surgical diagnosis. Three evaluations of MRA were separated by 3-week intervals: PS first; then transverse; then one image from each plane was made available *Ref. test interpretation*: Single specialised shoulder surgeon undertook surgery. Surgical report had to provide precise description of the subscapularis tendon before inclusion
Yagci, 2001[97] Radiology/orthopaedics Turkey (1997–8) Prosp2 Recruitment not reported Popl: NR patients/ index: 24 shoulders/ ref. test: 24 shoulders *Mean age*: 52 (16–73) *Sex*: 29% male *Outcomes*: Any; FT; PT Prevalence: 63; 42; 21%	Clinically suspected labral or rotator cuff abnormalities. All referred for surgery (therapeutic)	Evaluated MRI (see Appendix 9) and MRA *MRI unit*: 1.0 T; shoulder surface coil *MRI sequences*: T1-weighted; T2-weighted; T2-weighted GRE *MRI planes*: AX and PC; PC; PS *Contrast*: Gadopentetate dimeglumine. Shoulder exercised for an average of 15 minutes (range 10–20 minutes). *MRA sequences*: Spin-echo T1-weighted; fat-suppressed T1-weighted *MRA planes*: AX; PC and PS	*Ref. test*: Arthroscopy ($n = 1$) or open surgery ($n = 24$) *Test interval*: Between MRI and MRA, mean 11 days (range 0–27 days). Between MRA and surgery, 7days (range 1–23 days) *Further investigations*: Not reported *Diagnostic criteria*: Standard criteria established in recent literature were used (no references given) *Test interpreters*: Experienced musculoskeletal radiologists ($n = 2$) *Ref. test interpretation*: Not reported

Appendix 13

MRA: sensitivity and specificity results

Study	Methods[a]			Any tear			Full tear			Partial tear		
				Pr	Se (95% CI)	Sp (95% CI)	Pr	Se (95% CI)	Sp (95% CI)	Pr	Se (95% CI)	Sp (95% CI)
Binkert, 2001[103]	Fat	Gad 2 (n = 30)	R1	40	0.92 (0.65 to 0.99)	0.94 (0.74 to 0.99)						
			R2		1.00 (0.76 to 1.00)	1.00 (0.82 to 1.00)						
		Gad 4 (n = 26)	R1	42	0.91 (0.62 to 0.98)	0.87 (0.62 to 0.96)						
			R2		1.00 (0.74 to 1.00)	0.93 (0.70 to 0.99)						
		Ringer (n = 32)	R1	25	1.00 (0.68 to 1.00)	0.79 (0.60 to 0.91)						
			R2		1.00 (0.68 to 1.00)	0.88 (0.69 to 0.96)						
		All (n = 88)	R1[c]	35	0.94 (0.79 to 0.98)	0.86 (0.75 to 0.93)						
			R2		1.00 (0.89 to 1.00)	0.93 (0.83 to 0.97)						
Funke, 1996[101]	Con	Gad	[b]	68[c]	0.82 (0.59 to 0.94)	0.63 (0.31 to 0.86)	56	0.86 (0.60 to 0.96)	0.91 (0.62 to 0.98)	12[d]	0.67 (0.21 to 0.94)	0.91 (0.72 to 0.97)
	Fat				1.00 (0.82 to 1.00)	0.88 (0.53 to 0.98)		1.00 (0.78 to 1.00)	1.00 (0.74 to 1.00)		1.00 (0.44 to 1.00)	0.95 (0.78 to 0.99)
Hodler, 1992[90]	Con	Gad	[b]	47	0.71 (0.47 to 0.87)	0.84 (0.62 to 0.94)	11	1.00 (0.51 to 1.00)	0.88 (0.72 to 0.95)	36	0.46 (0.23 to 0.71)	0.96 (0.79 to 0.99)
Loew, 2000[102]	Con; 0.2 T	Gad	[b]				44	1.00 (0.76 to 1.00)	1.00 (0.80 to 1.00)			
	Fat; 1.5 T	Gad						1.00 (0.76 to 1.00)	1.00 (0.80 to 1.00)			
Pfirrmann, 1999[100]	Con, AC, PS, AX	Gad	R1[b]	42	0.90 (0.71 to 0.97)	0.86 (0.69 to 0.95)						
			R2		0.90 (0.71 to 0.97)	0.79 (0.62 to 0.90)						
	Con; PS	Gad	R1		0.95 (0.77 to 0.99)	0.55 (0.38 to 0.72)						
			R2		1.00 (0.85 to 1.00)	0.62 (0.44 to 0.77)						
	Con; AX	Gad	R1		0.90 (0.71 to 0.97)	0.76 (0.58 to 0.88)						
			R2		0.90 (0.71 to 0.97)	0.90 (0.74 to 0.96)						
Yagci, 2001[97]	Con	Gad	R1[b]	63	1.00 (0.80 to 1.00)	0.78 (0.45 to 0.94)	42	1.00 (0.72 to 1.00)	1.00 (0.78 to 1.00)	21	1.00 (0.57 to 1.00)	0.89 (0.69 to 0.97)
			R2		1.00 (0.80 to 1.00)	0.89 (0.56 to 0.98)		1.00 (0.72 to 1.00)	1.00 (0.78 to 1.00)		1.00 (0.57 to 1.00)	0.95 (0.75 to 0.99)

Pr, prevalence (%); Se, sensitivity; Sp, specificity.

[a] MRI details: Con, conventional MRI; Fat, fat-suppressed MRI; AX, axial view; AC, angled coronal; PS, parasaggital. Contrast: Gad, gadolinium contrast agent; Ringer, Ringer solution.

[b] Result used in meta-analysis (where more than one result per outcome, the most conservative was chosen).

[c] One partial tear detected at MRA but not at arthrography. If include this as true-positive instead of false-positive as above: prevalence = 0.72; Con, Se 0.83 (95% CI: 0.61 to 0.94), Sp 0.71 (95% CI: 0.36 to 0.92); Fat, Se 1.00 (95% CI: 0.82 to 1.00), Sp 1.00 (95% CI: 0.65 to 1.00).

[d] If include additional PT as true-positive instead of false-positive, as above: Pr = 0.16; Con, Se 0.50 (95% CI: 0.15 to 0.85), Sp 0.90 (95% CI: 0.71 to 0.97); Fat, Se 1.00 (95% CI: 0.51 to 1.00), Sp 1.00 (95% CI: .85 to 1.00).

Appendix 14

MRA: positive and negative likelihood ratios

Study	Methods[a]			Any tear			Full tear			Partial tear		
				Pr	LR+ (95% CI)	LR– (95% CI)	Pr	LR+ (95% CI)	LR– (95% CI)	Pr	LR+ (95% CI)	LR– (95% CI)
Binkert, 2001[103]	Fat	Gad 2 (n = 30)	R1	40	16.5 (2.4 to 111.7)	0.1 (0.0 to 0.6)						
			R2		36.5 (2.4 to 564.3)	0.0 (0.0 to 0.6)						
		Gad 4 (n = 26)	R1	42	6.8 (1.6 to 25.1)	0.1 (0.0 to 0.7)						
			R2		15.0 (2.3 to 99.6)	0.0 (0.0 to 0.7)						
		Ringer (n = 32)	R1	25	4.8 (2.2 to 10.5)	0.1 (0.0 to 1.1)						
			R2		8.0 (2.8 to 23.1)	0.1 (0.0 to 1.0)						
		All (n = 88)	R1[b]	35	6.7 (3.5 to 12.8)	0.1 (0.0 to 0.3)						
			R2		14.2 (5.5 to 36.7)	0.0 (0.0 to 0.3)						
Funke, 1996[101]	Con	Gad	[b]	68	2.2 (0.9 to 5.5)	0.3 (90.1 to 0.9)	56	9.4 (1.4 to 61.8)	0.2 (0.0 to 0.6)	12	7.3 (1.6 to 34.4)	0.4 (0.1 to 1.8)
	Fat				8.0 (1.3 to 50.0)	0.0 (0.0 to 0.5)		23.2 (1.5 to 350.5)	0.0 (0.0 to 0.5)		22.0 (3.2 to 149.3)	0.1 (0.0 to 1.8)
Hodler, 1992[90]	Con	Gad	[b]	47	4.5 (1.5 to 13.2)	0.3 (0.2 to 0.7)	11	8.0 (3.2 to 20.0)	0.1 (0.0 to 1.6)	36	10.6 (1.4 to 78.8)	0.6 (0.3 to 0.9)
Loew, 2000[102]	Con; 0.2 T	Gad	[b]				44	30.8 (2.0 to 471.9)	0.0 (0.3 to 0.9)			
	Fat; 1.5 T	Gad						30.8 (2.0 to 471.9)	0.0 (0.3 to 0.9)			
Pfirrmann, 1999[100]	Con, PS, AX	Gad	R1[b]	42	6.6 (2.6 to 16.5)	0.1 (0.0 to 0.4)						
			R2		4.4 (2.1 to 9.0)	0.1 (0.0 to 0.5)						
	Con; PS	Gad	R1		3.7 (1.9 to 7.3)	0.1 (0.0 to 0.6)						
			R2		8.7 (3.0 to 25.8)	0.0 (0.0 to 0.6)						
	Con; Ax	Gad	R1		2.1 (1.7 to 4.2)	0.1 (0.0 to 0.5)						
			R2		2.6 (1.7 to 4.2)	0.1 (0.0 to 0.4)						
Yagci, 200[197]	Con	Gad	R1[b]	63	4.5 (1.3 to 15.3)	0.0 (0.0 to 0.7)	42	28.6 (1.9 to 438.3)	0.0 (0.0 to 0.7)	21	9.5 (2.6 to 35.2)	0.1 (0.0 to 1.4)
			R2		9.0 (1.4 to 57.1)	0.0 (0.0 to 0.6)		28.6 (1.9 to 438.3)	0.0 (0.0 to 0.7)		19.0 (2.8 to 128.0)	0.1 (0.0 to 1.3)

Pr, prevalence (%); LR+, positive likelihood ratio; LR–, negative likelihood ratio.

[a] MRI details: Con, conventional MRI; Fat, fat-suppressed MRI; AX, axial view; PS, parasagittal. Contrast: Gad, gadolinium contrast agent; Ringer, Ringer solution.

[b] Result used in meta-analysis (where more than one result per outcome, the most conservative was chosen).

Appendix 15

Review of cost-effectiveness paper

Oh and colleagues, 1999[27]

Study question
The paper by Oh and colleagues presents a decision analytic model to consider the diagnostic choices in a patient with internal derangement of the shoulder and the cost-effectiveness of the diagnostic alternatives. The model compares estimates of the effectiveness and cost-effectiveness of conventional MRI and conventional arthography (double-contrast) with a hypothetical strategy of arthrogram (single-contrast) followed by MR arthrogram where indicated.

Setting
The paper is a USA study and the study setting is that of outpatient care for diagnostic procedures.

Approach
The study used a standard cost-effectiveness approach to estimate the cost per additional true diagnosis, although the authors do not explicitly state the unit of outcome used in the cost-effectiveness analysis (referring to accuracy only).

Model
The economic model used is a simple decision-tree model (based on Decisionmaker 7.05 software) involving three diagnostic strategies, conventional arthography, conventional MRI and MR arthrogram used as an adjunct to conventional arthography.

Data
Data on the prevalence of diagnostic outcomes (i.e. FT RCTs, PT RCTS, labral tears, absence of tear) are based on clinical experience. Effectiveness data were drawn from a review of English language publications from 1985 to 1997, with data taken from studies where patients had clinical symptoms suggesting shoulder impingement or instability. The authors present data on patient samples and the true positive rates from the selected studies.

Cost data are based on 1997 Medicare reimbursement rates (fee schedule for Pennsylvania, Metropolitan DC, New Jersey and Delaware providers). Imaging strategies are assumed to be undertaken on an outpatient basis only and costs are outpatient costs. The global imaging costs for conventional MRI, conventional arthrogram and MR arthrogram are reported as $464.57, $163.77 and $628.54, respectively. Following diagnosis for FT RCTs, PT RCTs and labral tears, patients were assumed to undergo repair procedures. Costs for surgical procedures for FT RCTs, PT RCTs and labral tears are detailed at $5411, $5324 and $4377 (Medicare rates). Costs for physical therapy and medication (NSAIDs) were not included in any of the treatment costs.

Effectiveness and cost data were from different sources.

Outcome/effectiveness
The outcome measure used, or measure of effectiveness, is the accuracy of the diagnostic strategy in the diagnosis of FT RCTs, PT RCTs, labral tears and the absence of a tear.

Utilities were discussed, but those referred to are not a measure of outcome or preference, or strength of preference. The utilities used can only be interpreted as an indication of the direction of benefit (e.g. +1.0 reflects a true-positive or true-negative outcome whereas –1.0 reflects a false-positive or false-negative outcome). However, it is difficult to see where the authors use these so-called utilities as effectiveness is based on findings for the accuracy of diagnosis. It may be that the accuracy data are transformed using the utilities; however, it is difficult to tell.

Discounted?
No discounting was done/reported.

Reported separately?
Cost and effect findings were presented separately for each of the three diagnostic strategies, average costs were used for each and the marginal cost of one strategy over another was calculated from these average costs.

Currency
US dollars.

Perspective
Not stated, but that of a healthcare purchaser was taken.

Statistical analysis?
No statistical analysis was undertaken.

Sensitivity analysis?
Undertaken on variables that were informed by expert opinion or those that were based on fewer than 10 scientifically proven cases.

Results of sensitivity analysis were presented for variations in effect of disease prevalence on the cost-effectiveness of each strategy. A range of scenarios were used, but the cost-effectiveness of MR arthrogram over conventional arthrogram was between $2669 and $4917, other than at the baseline assumption (with cost-effectiveness at $21,029).

Further sensitivity analysis was undertaken to evaluate the effect of uncertain true-positive rates used in the baseline analysis. The effect of independent changes in the true-positive rates of labral tears and partial tears for single-contrast (MR arthrogram) showed that MR arthrogram was more effective that MRI and arthrography in all variations. When the sensitivity of double-contrast arthrography (conventional arthrography) for labral tears and partial tears were independently increased, the marginal cost-effectiveness of MR arthrogram and MRI decreased owing to its greater cost.

Results
Base-case analysis presents findings for the average effectiveness of conventional arthrogram, conventional MRI and MRA (as adjunct therapy) to be 0.6610, 0.6715 and 0.7204, respectively. The average costs for these three diagnostic strategies are estimated to be $1090, $2033 and $2339, respectively. The authors present findings for the marginal cost-effectiveness of conventional MRI compared with conventional arthrogram, $89,895, MR arthrogram as adjunct therapy compared with conventional arthrogram, $21,029, and MR arthrogram compared with MRI, $6250. The cost-effectiveness results are interpreted as cost per accurate diagnosis (i.e. utility value of +1.0), although this is not explicitly stated by the authors.

Other
Computed tomography was not included as a strategy owing to its inability to diagnose labral tears and ultrasound was not included as a strategy owing to the varied and uncertain nature of its diagnostic properties.

The authors state that of the studies used to inform on the accuracy of the diagnostic strategies, many used patient populations that were not fully representative of the spectrum of patients on whom the tests might be used. The authors also highlight that the paucity of data on partial tears and the prevalence and clinical importance of partial tears limit some of the study conclusions.

The authors state that as the study did not use pure cost data, downstream costs (such as lost productivity and rehabilitation costs) were not evaluated.

It is difficult to understand the method for calculating the average cost of strategies (e.g. $1090 for arthrography) – obviously costs for diagnosis and interventions were weighted in the cost-effectiveness decision model, but the cost breakdowns/components are not presented.

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

Chair,
Professor Kent Woods,
Director, NHS HTA Programme & Professor of Therapeutics, University of Leicester

Professor Bruce Campbell, Consultant Vascular & General Surgeon, Royal Devon & Exeter Hospital

Professor Shah Ebrahim, Professor in Epidemiology of Ageing, University of Bristol

Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Radcliffe Hospital, Oxford

Dr Ron Zimmern, Director, Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

HTA Commissioning Board

Members

Programme Director,
Professor Kent Woods, Director, NHS HTA Programme, Department of Medicine and Therapeutics, Leicester Royal Infirmary, Robert Kilpatrick Clinical Sciences Building, Leicester

Chair,
Professor Shah Ebrahim, Professor in Epidemiology of Ageing, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol

Deputy Chair,
Professor Jenny Hewison, Professor of Health Care Psychology, Academic Unit of Psychiatry and Behavioural Sciences, University of Leeds School of Medicine, Leeds

Professor Douglas Altman, Professor of Statistics in Medicine, Centre for Statistics in Medicine, Oxford University, Institute of Health Sciences, Cancer Research UK Medical Statistics Group, Headington, Oxford

Professor John Bond, Professor of Health Services Research, Centre for Health Services Research, University of Newcastle, School of Health Sciences, Newcastle upon Tyne

Professor John Brazier, Director of Health Economics, Sheffield Health Economics Group, School of Health & Related Research, University of Sheffield, ScHARR Regent Court, Sheffield

Dr Andrew Briggs, Public Health Career Scientist, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Oxford

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Cloudside, Rossendale, Lancs
and
Principal Research Fellow, Clinical Therapeutics in the School of Pharmacy, Bradford University, Bradford

Professor Nicky Cullum, Director of Centre for Evidence Based Nursing, Department of Health Sciences, University of York, Research Section, Seebohm Rowntree Building, Heslington, York

Dr Andrew Farmer, Senior Lecturer in General Practice, Department of Primary Health Care, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Fiona J Gilbert, Professor of Radiology, Department of Radiology, University of Aberdeen, Lilian Sutton Building, Foresterhill, Aberdeen

Professor Adrian Grant, Director, Health Services Research Unit, University of Aberdeen, Drew Kay Wing, Polwarth Building, Foresterhill, Aberdeen

Professor Alastair Gray, Director, Health Economics Research Centre, University of Oxford, Institute of Health Sciences, Headington, Oxford

Professor Mark Haggard, Director, MRC ESS Team, CBU Elsworth House, Addenbrooke's Hospital, Cambridge

Professor F D Richard Hobbs, Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham, Primary Care and Clinical Sciences Building, Edgbaston, Birmingham

Professor Peter Jones, Head of Department, University Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Cambridge

Professor Sallie Lamb, Research Professor in Physiotherapy/Co-Director, Interdisciplinary Research Centre in Health, Coventry University, Coventry

Dr Donna Lamping, Senior Lecturer, Health Services Research Unit, Public Health and Policy, London School of Hygiene and Tropical Medicine, London

Professor David Neal, Professor of Surgical Oncology, Oncology Centre, Addenbrooke's Hospital, Cambridge

Professor Tim Peters, Professor of Primary Care Health Services Research, Division of Primary Health Care, University of Bristol, Cotham House, Cotham Hill, Bristol

Professor Ian Roberts, Professor of Epidemiology & Public Health, Intervention Research Unit, London School of Hygiene and Tropical Medicine, London

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital NHS Trust, Bramwell Dott Building, Edinburgh

Professor Martin Severs, Professor in Elderly Health Care, Portsmouth Institute of Medicine, Health & Social Care, St George's Building, Portsmouth

Dr Jonathan Shapiro, Senior Fellow, Health Services Management Centre, Park House, Birmingham

Current and past membership details of all HTA 'committees' are available from the HTA website (www.ncchta.org)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Dr Paul Cockcroft, Consultant Medical Microbiologist/ Laboratory Director, Public Health Laboratory, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, Addenbrooke's Hospital, Cambridge

Dr David Elliman, Consultant in Community Child Health, London

Dr Andrew Farmer, Senior Lecturer in General Practice, Institute of Health Sciences, University of Oxford

Dr Karen N Foster, Clinical Lecturer, Dept of General Practice & Primary Care, University of Aberdeen

Professor Jane Franklyn, Professor of Medicine, University of Birmingham

Professor Antony J Franks, Deputy Medical Director, The Leeds Teaching Hospitals NHS Trust

Mr Tam Fry, Honorary Chairman, Child Growth Foundation, London

Dr Susanne M Ludgate, Medical Director, Medical Devices Agency, London

Dr William Rosenberg, Senior Lecturer and Consultant in Medicine, University of Southampton

Dr Susan Schonfield, CPHM Specialised Services Commissioning, Croydon Primary Care Trust

Dr Margaret Somerville, Director of Public Health, Teignbridge Primary Care Trust, Devon

Mr Tony Tester, Chief Officer, South Bedfordshire Community Health Council, Luton

Dr Andrew Walker, Senior Lecturer in Health Economics, University of Glasgow

Professor Martin J Whittle, Head of Division of Reproductive & Child Health, University of Birmingham

Dr Dennis Wright, Consultant Biochemist & Clinical Director, Pathology & The Kennedy Galton Centre, Northwick Park & St Mark's Hospitals, Harrow

Pharmaceuticals Panel

Members

Chair,
Dr John Reynolds, Clinical Director, Acute General Medicine SDU, Oxford Radcliffe Hospital

Professor Tony Avery, Professor of Primary Health Care, University of Nottingham

Professor Iain T Cameron, Professor of Obstetrics & Gynaecology, University of Southampton

Mr Peter Cardy, Chief Executive, Macmillan Cancer Relief, London

Dr Christopher Cates, GP and Cochrane Editor, Bushey Health Centre, Bushey, Herts.

Mr Charles Dobson, Special Projects Adviser, Department of Health

Dr Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Dr Karen A Fitzgerald, Pharmaceutical Adviser, Bro Taf Health Authority, Cardiff

Professor Alastair Gray, Professor of Health Economics, Institute of Health Sciences, University of Oxford

Mrs Sharon Hart, Managing Editor, *Drug & Therapeutics Bulletin*, London

Dr Christine Hine, Consultant in Public Health Medicine, Bristol South & West Primary Care Trust

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Dr Frances Rotblat, CPMP Delegate, Medicines Control Agency, London

Mrs Katrina Simister, New Products Manager, National Prescribing Centre, Liverpool

Dr Ken Stein, Senior Lecturer in Public Health, University of Exeter

Professor Terence Stephenson, Professor of Child Health, University of Nottingham

Dr Richard Tiner, Medical Director, Association of the British Pharmaceutical Industry, London

Professor Dame Jenifer Wilson-Barnett, Head of Florence Nightingale School of Nursing & Midwifery, King's College, London

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell, Consultant Vascular and General Surgeon, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Head of Clinical Support & Health Protection, Directorate of Health and Social Care (North), Department of Health, Manchester

Professor John Bond, Head of Centre for Health Services Research, University of Newcastle upon Tyne

Mr Michael Clancy, Consultant in A & E Medicine, Southampton General Hospital

Dr Carl E Counsell, Senior Lecturer in Neurology, University of Aberdeen

Dr Keith Dodd, Consultant Paediatrician, Derbyshire Children's Hospital, Derby

Professor Gene Feder, Professor of Primary Care R&D, Barts & the London, Queen Mary's School of Medicine and Dentistry, University of London

Ms Bec Hanley, Freelance Consumer Advocate, Hurstpierpoint, West Sussex

Professor Alan Horwich, Director of Clinical R&D, The Institute of Cancer Research, London

Dr Phillip Leech, Principal Medical Officer for Primary Care, Department of Health, London

Mr George Levvy, Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay, Professor of Psychiatry for the Elderly, University of Leicester

Dr Mike McGovern, Senior Medical Officer, Heart Team, Department of Health, London

Dr John C Pounsford, Consultant Physician, North Bristol NHS Trust

Professor Mark Sculpher, Professor of Health Economics, Institute for Research in the Social Services, University of York

Dr L David Smith, Consultant Cardiologist, Royal Devon & Exeter Hospital

Professor Norman Waugh, Professor of Public Health, University of Aberdeen

Expert Advisory Network

Members

Mr Gordon Aylward,
Chief Executive,
Association of British Health-Care Industries, London

Ms Judith Brodie,
Head of Cancer Support Service, Cancer BACUP, London

Mr Shaun Brogan,
Chief Executive, Ridgeway Primary Care Group, Aylesbury, Bucks

Ms Tracy Bury,
Project Manager, World Confederation for Physical Therapy, London

Mr John A Cairns,
Professor of Health Economics, Health Economics Research Unit, University of Aberdeen

Professor Howard Stephen Cuckle,
Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Professor Nicky Cullum,
Director of Centre for Evidence Based Nursing, University of York

Dr Katherine Darton,
Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux,
Professor of Paediatric Epidemiology, London

Professor Martin Eccles,
Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby,
Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Mr Leonard R Fenwick,
Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Professor David Field,
Professor of Neonatal Medicine, Child Health, The Leicester Royal Infirmary NHS Trust

Mrs Gillian Fletcher,
Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield, West Sussex

Ms Grace Gibbs,
Deputy Chief Executive, Director for Nursing, Midwifery & Clinical Support Servs., West Middlesex University Hospital, Isleworth, Middlesex

Dr Neville Goodman,
Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Robert E Hawkins,
CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor F D Richard Hobbs,
Professor of Primary Care & General Practice, Department of Primary Care & General Practice, University of Birmingham

Professor Allen Hutchinson,
Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Professor Rajan Madhok,
Medical Director & Director of Public Health, Directorate of Clinical Strategy & Public Health, North & East Yorkshire & Northern Lincolnshire Health Authority, York

Professor David Mant,
Professor of General Practice, Department of Primary Care, University of Oxford

Professor Alexander Markham,
Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Dr Chris McCall,
General Practitioner, The Hadleigh Practice, Castle Mullen, Dorset

Professor Alistair McGuire,
Professor of Health Economics, London School of Economics

Dr Peter Moore,
Freelance Science Writer, Ashtead, Surrey

Dr Andrew Mortimore,
Consultant in Public Health Medicine, Southampton City Primary Care Trust

Dr Sue Moss,
Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, Surrey

Professor Jon Nicholl,
Director of Medical Care Research Unit, School of Health and Related Research, University of Sheffield

Mrs Julietta Patnick,
National Co-ordinator, NHS Cancer Screening Programmes, Sheffield

Professor Chris Price,
Visiting Chair – Oxford, Clinical Research, Bayer Diagnostics Europe, Cirencester

Ms Marianne Rigge,
Director, College of Health, London

Professor Sarah Stewart-Brown,
Director HSRU/Honorary Consultant in PH Medicine, Department of Public Health, University of Oxford

Professor Ala Szczepura,
Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor,
Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster,
Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.ncchta.org) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org

ISSN 1366-5278