The validity and accuracy of clinical tests used to detect labral pathology of the shoulder: a systematic review

Munro W, Healy R

CRD summary
This review found limited evidence from single studies that Biceps Load tests I and II, IRRT, Kim test and Jerk test were accurate in differentiating labral pathology. Subgroup analysis showed varying results for Crank and Active Compression tests. Potential for missed studies and limitations in the analysis mean that these conclusions are unlikely to be reliable.

Authors' objectives
To evaluate the accuracy of clinical tests to detect labral pathology of the shoulder.

Searching
MEDLINE, CINAHL, AMED, The Cochrane Library, DARE and HTA Database were searched from inception to June 2007 for published English-language studies. Search terms were reported and included a diagnostic filter. References of retrieved studies were screened and a citation search was conducted.

Study selection
Diagnostic cohort or case-control studies that evaluated clinical examination tests to detect labral pathology in patients with shoulder pain in comparison to a reference standard and that reported data on sensitivity and specificity were eligible for inclusion. Patients with other pathologies leading to shoulder pain were excluded.

Included studies evaluated the following clinical tests: abduction inferior stability; active compression; anterior slide; biceps load I and II; crank; clunk; compression rotation; forced abduction; internal rotation resistance test (IRRT); new pain provocation; posterior impingement; resisted supination external rotation; resistance test; jerk; Kim; and posterior jerk. Most included studies used arthroscopy as reference standard; other reference standards used were magnetic resonance imaging (MRI), radiography, surgical findings and clinical data. Mean age ranged from 21 to 46 years. All studies were conducted in specialist settings among patients referred to hospital consultants for surgery.

One reviewer assessed studies for inclusion. In the case of uncertainty a second reviewer was consulted.

Assessment of study quality
Two reviewers independently assessed study quality using QUADAS.

Data extraction
Two reviewers independently extracted data as 2x2 tables of test performance; 0.5 was added to all cells to account for 0 values. Sensitivity, specificity, positive likelihood ratios (LR+) and negative (LR-) likelihood ratios were calculated for each study. Disagreements were resolved through consensus.

Methods of synthesis
Sensitivity and specificity were plotted in summary receiver operating characteristic (ROC) space and summary ROC (SROC) curves were estimated for tests evaluated in more than two studies. Positive and negative likelihood ratios were plotted on forest plots.

Results of the review
Fifteen studies (n=1,919) were included in the review; only 11 of these reported sufficient information to contribute to the data analysis. Studies suffered from a number of methodological limitations related to spectrum bias, selection bias, verification bias, clinical review bias, diagnostic review bias, reporting of reference test details, availability of clinical information and information regarding disease progression between the tests. Nine studies suffered from two or fewer key sources of bias, four suffered from two to four key sources of bias and three suffered from at least five
key sources of bias (it was unclear why these totalled 16 rather than 15 studies). 

Most clinical tests were each evaluated in only one study. The ROC plot showed that most studies reported values towards the upper left hand corner with values of specificity above 80% and sensitivity above 70%. However, five sets of results showed much poorer accuracy with estimates of sensitivity or specificity less than 50%. Five tests evaluated in single studies that suffered from two or less key sources of bias were found to have both high and sensitivity and specificity based on studies of moderately sound methodological quality: Biceps Load test I; Biceps Load test II; IRRT; Kim test; and Jerk test.

Only two tests were evaluated by more than two studies and reported in sufficient detail to allow extraction of 2x2 data. Crank test (n=3) showed substantial variation: two studies suggested good accuracy (sensitivity and specificity both above 83%); and one study suggested much poorer accuracy (sensitivity of 46% and specificity of 56%). Crank test was also evaluated in four studies that did not contribute to the analysis; sensitivity ranged from 9% to 58% and specificity ranged from 70% to 83% in these studies. Active compression test (n=3) also showed substantial variation: one study suggested 100% sensitivity and 99% specificity; and two studies reported values of sensitivity that ranged from 48% to 54% and specificity that ranged from 31% to 55%. This test was also evaluated in four studies that did not contribute to the analysis (sensitivity ranged from 54% to 65% and specificity ranged from 11% to 73%).

Authors' conclusions
There was limited evidence from single well-carried out tests that Biceps Load tests I and II, IRRT, Kim test and Jerk test were accurate in differentiating labral pathology from other pathologies in selected populations. Subgroup analysis showed varying results for Crank test and Active Compression test.

CRD commentary
The review addressed a broad question with defined inclusion criteria. An appropriate range of databases was searched for published studies. Use of a diagnostic filter and restriction of the review to published English-language studies meant that it was likely that relevant studies were missed. Appropriate steps were taken to minimise risk of bias and errors in the extraction of studies and quality assessment; no such steps were taken when selecting studies, which raised the possibility of missed studies. Study quality was assessed using appropriate criteria and these were considered in the analysis of results. Studies were judged to have been high quality if they suffered from two or fewer key sources of bias. However, studies that suffered from even one key source of bias could produce biased results. It would have been more informative to investigate the effects of different sources of bias on results rather than grouping studies according to the number of key sources of biases present. The review lacked details of the included studies (especially in relation to participants) and so generalisability of the review findings was unclear. Summarising results using SROC plots provided a helpful overview of the results of the review. Most of the results of the review were summarised narratively; however, results were pooled for the two tests evaluated in more than two studies. Results from these studies were very heterogeneous and use of pooling under these circumstances was questionable. It was unclear why studies were excluded from the narrative analysis if they did not report sufficient data to extract a 2x2 table as they still provided data on sensitivity and specificity that could have been discussed. Differences between studies were not adequately investigated. The authors acknowledge that their conclusions were based on single studies and this, combined with methodological limitations of the review in terms of potential for missed studies and limitations in the analysis, means that the conclusions are unlikely to be reliable.

Implications of the review for practice and research
Practice: The authors did not state any implications for practice.

Research: The authors stated that there was a need for further evaluation of labral pathology tests in different populations by less skilled examiners. Such studies should address key sources of bias for diagnostic tests and provide more details of demographic information and adequate raw data.

Funding
Not stated.
Bibliographic details

PubMedID
18996735

DOI
10.1016/j.math.2008.08.008

Original Paper URL
http://www.manualtherapyjournal.com/article/S1356-689X(08)00142-2/abstract

Indexing Status
Subject indexing assigned by NLM

MeSH
Cartilage, Articular /pathology; Diagnostic Imaging /methods /standards; Female; Humans; Joint Instability /diagnosis; Magnetic Resonance Imaging /methods; Male; Physical Examination /methods /standards; Reproducibility of Results; Rotator Cuff /injuries /pathology; Sensitivity and Specificity; Shoulder Impingement Syndrome /diagnosis; Shoulder Joint /pathology /physiopathology; Shoulder Pain /diagnosis

AccessionNumber
12009106085

Date bibliographic record published
30/09/2009

Date abstract record published
19/05/2010

Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.