Available online at www.sciencedirect.com
ScienceDirect
Manual Therapy 14 (2009) 119–130
MANUAL THERAPY
www.elsevier.com/math

Systematic Review

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

Wendy Munro*, Raymond Healy

University of Salford, Salford, Greater Manchester M6 6PU, UK

Received 16 July 2007; received in revised form 8 August 2008; accepted 27 August 2008

Abstract

Labral tears frequently require repair [Kim S, Ha K, Han K. Biceps Load test: a clinical test for superior labrum anterior and posterior lesions in shoulders with recurrent anterior dislocations. The American Journal of Sports Medicine 1999;27(3):300–3]. Physiotherapists need confidence in clinical tests used to detect labral pathology to accurately identify this condition. This review systematically evaluates the evidence for the accuracy of these tests with reference to study quality and key biases.

Cochrane, Medline, Cinahl, AMED, DARE and HTA databases were searched to identify 15 studies evaluating 15 clinical tests for labral pathology against Magnetic Resonance Imaging MRI or surgery. Two independent reviewers assessed methodological quality using Quality Assessment of Diagnostic Accuracy Studies (QUADAS). Meta Disc calculated likelihood ratios (positive LR > 10, providing convincing diagnostic evidence of ruling a condition in; negative LR < 0.2 providing large to moderate evidence of ruling the condition out) and true positive rates (TPRs) against false positive rates (FPRs) in receiver operator characteristic (ROC) plots and summary receiver operator curves (SROCs).

Probable overestimation of accuracy was caused by use of case control design, verification bias and use of a lesser reference standard. Six accurate tests; Biceps Load I (+LR: 29.09; −LR: 0.09) Biceps Load II (+LR: 26.32; −LR: 0.11), Internal Rotation Resistance (IRRT) (+LR: 24.77; −LR: 0.12), Crank (+LR: 13.59 and 6.46; −LR: 0.1 and 0.22), Kim (+LR: 12.62; −LR:0.21) and Jerk (+LR: 34.71; −LR: 0.27) tests were identified from high quality single studies in selected populations. Subgroup analysis identified varying results of accuracy in the Crank test and the Active Compression (AC) test when evaluated in more than one study. Further evaluation is needed before these tests can be used with confidence.

Keywords: Labral pathology; Screening; Sensitivity and specificity; Likelihood ratios

1. Introduction

Assessment and diagnosis has become an increasingly important aspect of the physiotherapist's role in clinical specialist and extended scope roles. Differential diagnosis of the shoulder is a problematic area, with no standardised definitions and diagnostic criteria for defining disorders being inconsistent and unreliable (Green et al., 2003). Hanchard et al. (2004) advocate an evidence based conservative management approach which does not differentiate between subacromial impingement syndrome (SIS), posterior superior glenoid impingement (PSGI) and superior labral anterior posterior (SLAP) lesions suggesting that such clear cut diagnosis is unnecessary. However, the presence of signs, possibly indicating glenoid labral damage e.g. pain on overhead activities, deep shoulder pain, painful catching and popping or clicking (Musgrave and Rodosky, 2001), should lead the clinician to consider further management outside the scope of physiotherapy such as arthroscopy or surgery. Symptoms of labral pathology can make it difficult to differentiate from other shoulder pathologies such as impingement and

* Corresponding author. Directorate of Physiotherapy, Mary Seacole Building, Frederick Road Campus, University of Salford, Salford, Greater Manchester M6 6PU, UK. Tel.: +44 0161 295 2502; fax: +44 0161 295 2432.
E-mail address: w.munro@salford.ac.uk (W. Munro).

1356-689X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.math.2008.08.008

acromio-clavicular joint arthritis (Musgrave and Rodosky, 2001). Knowledge of the tests available to assist in the differentiation of this diagnosis, the validity of these tests and the skills to perform them are therefore required. Physical examination has been described as more of an art than a science although carefully planned diagnostic test accuracy studies will provide more of a science to this art (Reider, 2004).

Although SLAP lesions commonly occur in the young active overhead athlete (Andrews et al., 1985) and following a compressive or distraction force on the shoulder (Andrews et al., 1985; Snyder et al., 1990; Maffet et al., 1995), labral pathology may result from a sudden fall onto the outstretched hand or elbow with the shoulder in a somewhat adducted and extended position. This can lead to secondary symptoms of impingement caused by superior translation of the humeral head (Kumar et al., 1989; Altchek et al., 1992; Schmitz, 1999). Hasan (2006) has suggested the superior labrum to have a more meniscoid attachment to the glenoid than the rest of the labrum, making it susceptible to degenerative as well as traumatic lesions. Tests for labral pathology therefore need to be accurate in both general and athletic population settings in a wide age group of patients.

Liume et al. (2004) and Jones and Galluch (2007) have systematically reviewed studies relating to clinical tests for instability and labral lesions and superior glenoid labral lesions respectively. Liume et al. (2004) reviewed 17 studies evaluating clinical tests for shoulder instability or labral lesion suggesting the Relocation test and the Anterior Release test to be most clinically relevant in diagnosing instability, and the Biceps Load tests I and II, the Pain provocation test and the Internal Rotation Resistance test (IRRT) to be most promising for labral tears. Jones and Galluch (2007) reviewed 12 studies and concluded that SLAP specific physical examination results cannot be used alone to diagnose SLAP lesions.

This review, including additional studies, focuses on studies evaluating tests for labral pathology and adds to the previous literature with a thorough quality assessment of the included studies using Quality Assessment of Diagnostic Accuracy Studies (QUADAS), receiver operating characteristic and forest plots. Previous studies have either used QUADAS only, or levels of evidence to control for study quality. Subgroup analysis is carried out on single tests evaluated in different studies.

2. Methods

2.1. Search strategy

Publications were identified by searching the following databases: Cochrane (1995–2007), Medline (1996–June 2007), Cinahl (1982–June 2007), AMED (1985–June 2007), Health Technology Assessment (1995–June 2007) and the Database of Abstracts of Reviews of Effectiveness (1995–June 2007). A combination of MeSh terms (exp 'sensitivity and specificity'/, exp shoulder joint/, exp joint instability/, exp shoulder injuries/, exp shoulder pain/) and text words (specificity, false negative, accuracy, screening, labral pathology, SLAP lesions, SLAP, glenoid labrum, instability and individual test names) based on Devillé et al.'s (2000) optimal search strategy were used. The search was limited to articles of English language.

2.2. Inclusion and exclusion criteria

The titles of the articles were screened and filtered and the abstracts of the filtered articles were screened by one reviewer (WM) for fulfilment of the inclusion and exclusion criteria. Inclusion criteria were: cohort and case control design, shoulder pain, clinical examination tests used to evaluate labral pathology, comparison against a reference standard, and inclusion of sensitivity and specificity values. Exclusions were: other pathologies leading to shoulder pain (e.g. referred from spine or internal organs, Cerebrovascular accident CVA) and studies omitting values of either sensitivity or specificity.

Where the first reviewer was uncertain whether a study should be included, a second reviewer (RH) was consulted and a decision made by consensus. To ensure completeness of the literature search, the references of the included studies were hand searched for further references and a citation search was carried out. No further studies were identified.

2.3. Data extraction and quality assessment

A standardised extraction form was piloted and then used independently by two reviewers (WM and RH) to maintain quality and objectivity (Deeks and Morris, 1996). Any disagreements were decided by consensus.

Quality assessment was carried out on all studies which met the inclusion and exclusion criteria using the QUADAS tool (Whiting et al., 2003). This ensured that all studies were evaluated for individual quality items rather than being given a quality score as advocated by Whiting et al. (2005). The QUADAS tool has been developed based on expert consensus and empirical evidence. It has been shown to have varied reliability, with agreement on individual checklist items of 90% (Whiting et al., 2006) 76% (Davis et al., 2007) and 78% (Hollingworth et al., 2006) and kappa scores of 0.65 (Whiting et al., 2006), 0.39 (Davis et al., 2007) and 0.22 (Hollingworth et al., 2006) demonstrating good, fair and fair inter-rater reliability respectively (Altman, 1999). Differences appear to be down to numbers of reviewers, working proximity of the reviewers and experience in diagnostic accuracy systematic reviews (Hollingworth et al., 2006). The QUADAS tool has gained positive feedback in a pilot study by twenty reviewers, with eighteen considering the tool to cover all important items (Whiting et al., 2006). The tool includes questions relevant to spectrum bias, selection bias, disease progression bias, verification bias, incorporation bias, execution of the index and reference test, index test and reference standard test review bias, uninterpretable tests and withdrawals from the studies. These terms are explained in Table 1. Each item was scored yes, no or unclear according to the scoring guidelines of the tool (Whiting et al., 2003). The meaning of the questions relevant to the clinical applicability was discussed and agreed by the reviewers prior to use of the tool. For the purpose of the review, it was assumed that