Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review

May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M

CRD summary
This review concluded that there was no consistent evidence that any physical examination procedure for muscular-related shoulder pain had acceptable levels of reliability or reproducibility. Variation between the included studies and their poor methodology and inconsistent findings mean that the authors’ cautious conclusions appear to be valid, despite the risk of missing data.

Authors’ objectives
To assess the reliability of physical examination procedures for the clinical examination of patients with shoulder pain.

Searching
MEDLINE, PEDro, AMED, PsycINFO, The Cochrane Library and CINAHL were searched for full reports of studies published in English up to June 2009. Search terms were reported. Reference lists of retrieved articles were screened for further studies.

Study selection
Studies that evaluated intra- and/or inter-examiner reliability of physical procedures for the examination of patients with any sort of muscular shoulder pain were eligible for inclusion in the review. Studies were excluded if they involved a mechanical device (except simple tape measures). Studies of only asymptomatic volunteers were excluded, but studies that included a mix of symptomatic and asymptomatic participants were included.

The most commonly evaluated physical examination techniques in the included studies assessed: impingement/rotator cuff; supraspinatus; infraspinatus; subscapularis; biceps; resisted tests; scapular positioning; scapular movement dysfunction; range of movement and pain response; end feel; accessory movements; movement diagrams; acromioclavicular joint; shoulder instability; glenoid labral tears; pectoralis minor length test; tenderness; and myofascial trigger points. Further details of the specific tests and items evaluated were reported in the review. Twenty-four of the included studies involved physical therapists alone (included students in one study), six included medical professionals alone and six included medical and other health professionals. Just over half of the studies investigated intertester reliability (30 studies), 16 investigated intra-tester reliability and nine studies investigated both parameters.

One reviewer initially screened studies for inclusion and discarded clearly irrelevant studies. The remaining studies were reviewed by two independent reviewers. Disagreements were resolved by consensus or a third reviewer if necessary.

Assessment of study quality
Two reviewers independently assessed the methodological quality of the studies using the criteria: adequate description of study population; patients selected consecutively or randomly; representative of clinical practice; sample size or sample calculation; clear description of study methods; consistency of procedures; adequate attempts to reduce bias; adequate description of examiners; consensus procedure prior to testing or pilot study; multiple examiners tested; multiple testing between examiners; standardised outcome measures; frequency of outcome and agreement reported; and reporting of appropriate inferential statistics and measure of variance. Studies were awarded a score between a maximum of 4 and 10 for each criterion (further details reported in the review) and a total score was reported for each study. Studies that scored at least 60% were considered to be of high quality. Disagreements were resolved through consensus.

Data extraction
Kappa or weighted kappa excepts were extracted, where reported, along with 95% confidence intervals (CIs) and intraclass correlation coefficients.
Two reviewers independently extracted the study data; disagreements were resolved through consensus.

**Methods of synthesis**
Studies were grouped by anatomical and pathological entities or by physical examination technique and inter-tester reliability. Results were reported in tables with a narrative summary. Agreement between the results was assessed using an intra-class correlation coefficient (ICC); coefficients of 0.85 (arbitrary cutoff) and above were considered to be reliable.

**Results of the review**
Thirty-six studies were included in the review (number of participants unknown). The mean methodological score for the studies was 57%. Total methodological scores ranged from 25% to 85%. Seventeen studies were deemed high quality (scored at least 60%).

There were inconsistent findings across all procedures from both high- and low-quality studies. Intra-tester results showed that 25 out of 54 investigations had values greater than 0.85. A further 11 had values that ranged from 0.70 to 0.85. These results tended to be slightly better than inter-tester results, which were very inconsistent (further details reported in the review). High-quality studies were less likely to meet the pre-agreed level of reliability. Most studies indicated poor reliability for all procedures investigated.

**Authors’ conclusions**
There was no consistent evidence that any physical examination procedure for shoulder pain had acceptable levels of reliability or reproducibility.

**CRD commentary**
This generally well-conducted review answered a clearly defined research question, but may have missed relevant data by limiting study inclusion to only full-study reports published in English. Little information was presented about the clinical characteristics of the included studies. Methodological quality of the studies was assessed and reported. There appeared to be a large variation in study quality. Most were judged not to be high quality, which suggested that the data may not have been reliable. The studies appeared to vary greatly with respect to the type of physical procedure assessed.

Variation between the included studies and their poor methodology and inconsistent findings mean that the authors’ cautious conclusions appear to be valid, despite the risk of missing data.

**Implications of the review for practice and research**
**Practice:** The authors stated that alternative and reliable methods for classifying muscular shoulder pain should be used in preference to physical examination techniques. Poor validity and reliability precluded physical examination techniques from making clinical diagnoses and should be abandoned in conservative care. Subgroups should be identified using reliable clinical characteristics.

**Research:** The authors stated that further studies were required to determine whether physical examination techniques may be more reliable when used in conjunction with other tests for the assessment of patients with muscular-related shoulder pain. Further research should aim to identify clinical characteristics with management and prognostic implications.

**Funding**
Not stated.

**Bibliographic details**
May S, Chance-Larsen K, Littlewood C, Lomas D, Saad M. Reliability of physical examination tests used in the assessment of patients with shoulder problems: a systematic review. Physiotherapy 2010; 96(3): 179-190
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.