A meta-analysis examining clinical test utilities for assessing meniscal injury

Meserve B B, Cleland J A, Boucher T R

CRD summary
This review assessed the accuracy of clinical tests for identifying meniscal lesions of the knee. It suffered from a number of methodological limitations that were acknowledged by the author. The conclusion that all clinical tests showed low to moderate accuracy and that more and higher quality research is needed to fully assess the accuracy of these tests is a reasonable interpretation of the available data.

Authors' objectives
To determine the accuracy of tests for assessing meniscal lesions of the knee.

Searching
MEDLINE and CINAHL were searched from inception to 2006. The Cochrane Library and Clinical Pharmacology, Health Source and Nursing/Academic Edition, and Sport Discus (1975-2006) were also searched. Search terms were reported. Searches were restricted to English language articles.

Study selection
Studies comparing at least one clinical test for identifying meniscal lesions of the knee with a reference standard (arthrotomy or arthroscopy) and reporting data on diagnostic accuracy were eligible for inclusion. Studies were excluded if the reference standard was not applied to all participants or the physical examination was conducted under anaesthesia. Study selection was performed by one reviewer. Studies included in the meta-analyses assessed the accuracy of Apley's test, McMurray's test and the joint line tenderness test; other included studies assessed pain on forced flexion, blocked extension, Ege's test, Thessaly test, modified pivot shift test and medial-lateral grind test. All included studies assessed both medial and lateral meniscus. Most studies were in majority male populations. Participant ages ranged from 10 to 78 years.

Assessment of study quality
Study quality was assessed by one reviewer using a 16-item checklist with a maximum score of 30 and variable weight between items, adapted from the Cochrane Methods Group on Systematic Review of Screening and Diagnostic Tests checklist. The checklist was reported in full in an appendix to the article and included items relating to study population, description of index test and reference standard and blinding of those interpreting test results. Only studies with a quality score of at least 50 per cent were included in the meta-analysis. Included studies were also assigned an evidence level (1 to 4), definitions of which were reported in the article.

Data extraction
One reviewer extracted data to populate 2 x 2 contingency tables (numbers of true positives, false negatives, false positives and true negatives) from each study or, where these data were unavailable, reported sensitivity and specificity values and the number of study participants. Sensitivity, specificity, positive and negative likelihood ratios, and diagnostic odds ratios (DOR) were the main outcome measures reported.

Methods of synthesis
Studies included in the meta-analysis were assigned a grouping based on quality score (<20, 20, or >20). Separate analyses of the variance of logit sensitivity and logit specificity were run against these score categories and score category was found to be statistically significant for both. Quality-corrected logit sensitivity and logit specificity values were therefore used to generate summary ROC curves for each test.

Results of the review
Eleven studies (nine cohort and two case-control) with a total of 1,768 participants were included in the review.

The quality score and level of evidence associated with each study included in the review was reported in the article; all included studies had quality scores of 50 per cent or more.
The pooled estimates of sensitivity, specificity and DOR (three studies, 731 participants) for Apley's test were: 0.22 (95% CI: 0.17, 0.28) for sensitivity, 0.88 (95% CI: 0.72, 0.96) for specificity and 2.20 (95% CI: 0.27, 17.66) for DOR.

The pooled estimates of sensitivity, specificity and DOR (eight studies, 1,374 participants) for the joint line tenderness test were 0.76 (95% CI: 0.73, 0.80) for sensitivity, 0.77 (95% CI: 0.64, 0.87) for specificity and 10.98 (95% CI: 3.02, 39.95) for DOR.

The pooled estimates of sensitivity, specificity and DOR (eight studies, 1,475 participants) for McMurray's test were 0.55 (95% CI: 0.50, 0.60) for sensitivity, 0.77 (95% CI: 0.62, 0.87) for specificity and 3.99 (95% CI: 1.04, 15.31) for DOR.

Accuracy data for all tests assessed in individual included studies were reported in full in the article.

Summary ROC curves and pooled estimates of accuracy showed joint line tenderness to be the best test, followed by McMurray's test then Apley's test. Thessaly's test appeared to have the best overall accuracy, but data for this test were derived from a single case-control study.

**Authors' conclusions**

The clinical tests studied demonstrated low to moderate accuracy. Future studies should include larger samples of participants without meniscal tears in order to improve estimates of specificity.

**CRD commentary**

The review addressed a clearly stated research question defined by appropriate inclusion criteria. Extensive literature searches were conducted and reported in full, but restriction to studies published in English may have resulted in loss of data. The review was conducted by a single investigator, leaving it potentially open to error and/or bias. The methodological quality of included studies was assessed, but the results of this assessment were reported as summary scores only. Summary quality scores were incorporated in the meta-analyses. Summary scores are not generally considered to be a useful way of investigating the effects of study quality upon diagnostic accuracy. Ideally, this should be done by including variables for individual quality items in the summary ROC model. It was unclear whether significant between study heterogeneity remained after the correction for study quality was applied and, therefore, whether the decision to generate pooled estimates of sensitivity, specificity and DOR was appropriate. The authors' conclusions were cautious and focused upon the need for further research; an appropriate interpretation of the data presented.

**Implications of the review for practice and research**

Practice: joint line tenderness was a better test than either McMurray's or Apley's. Ege's and Thessaly's tests had the highest accuracy, but from smaller samples derived from single studies.

Research: future studies should be of higher methodological quality and include larger samples of participants without meniscal tears in order to improve estimates of specificity. In addition, studies should evaluate multiple clinical tests in the same participant group, to allow direct comparisons of accuracy.

**Funding**

Not reported

**Bibliographic details**


**PubMedID**

18212035

**DOI**
Indexing Status
Subject indexing assigned by NLM

MeSH
Evidence-Based Medicine; Humans; Knee Injuries /diagnosis; Magnetic Resonance Imaging; Menisci, Tibial /injuries; Physical Examination /methods; ROC Curve; Sensitivity and Specificity

AccessionNumber
12008102655

Date bibliographic record published
01/09/2008

Date abstract record published
31/03/2009

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.