

Validity of the McMurray's Test and Modified Versions of the Test: A Systematic Literature Review

WAYNE HING, PhD, MSc, ADP, Dip MT, Dip Phys, MNZCP¹; STEVE WHITE, MHSc, DipPH, DipMT, DipPhys²;
DUNCAN REID, MHSc, PGDHSc, Dip MT, Dip Phys, BSc³; ROB MARSHALL, BSc, PGDHSc⁴

A wide variety of clinical tests are used to diagnose meniscal pathology within the knee joint. Palpation for joint line tenderness, the Apley's Grind test, and the McMurray's test are commonly used in physical therapy practice¹. The accurate diagnosis of meniscal pathology on the basis of the

findings of such tests is often difficult. A recent evidence-based guideline for the management of acute soft tissue injuries to the knee has recommended that joint line tenderness is the only reliable clinical indicator of meniscal pathology². The possibility of there being associated intra-articular pathology (such as anterior

cruciate ligament rupture) confounds results, and the unknown validity, sensitivity, and specificity of the tests make it difficult for the clinician to be confident in making a definitive diagnosis³.

The McMurray's test, as described in Corea et al⁴, was designed to detect tears in the posterior segment of the meniscus. It is performed by placing the knee beyond 90° of flexion and then rotating the tibia on the femur into full internal rotation to test the lateral meniscus, or full external rotation to test the medial meniscus. The same maneuvers are performed in gradually increasing degrees of knee flexion to progressively load more posterior segments of the menisci. No valgus or varus stress is applied. During the maneuver, the joint line is palpated both medially and laterally. A positive test is considered to be a *thud* or *click* that *can sometimes be heard but can always be felt*⁴ (Figure 1).

The findings of studies testing the validity of the McMurray's test have varied widely, mostly due to variations in the size and type of the study population as well as differences in description and application of the test³. More recent research has shown that modifications to the original McMurray's test may have better validity and diagnostic accuracy than the original McMurray's test^{3,5-8}. The

ABSTRACT: Clinical assessment of meniscal pathology in the knee has proven difficult due to the wide number of tests available and variations in their interpretation and application. The purpose of this paper was to assess the literature investigating the validity and diagnostic accuracy of the McMurray's test (and modifications) for determining meniscal pathology of the knee so that conclusions could be drawn regarding its clinical usefulness as a test. Electronic databases (Medline, CINHALL, AMED, SPORTSDiscus, and SCOPUS) were searched from March 1980 to May 2008. In addition, cited references of relevant articles were examined. Studies were included for analysis if they compared the McMurray's test with a gold standard of knee arthroscopy or magnetic resonance imaging (MRI). Eleven studies met the inclusion criteria. Collectively, these studies indicate that there is little consensus in the reported measures of validity of the McMurray's test and that this is mostly due to limitations in the methodological quality of the studies that were assessed. Methodological scores on the STARD (Standards for Reporting of Diagnostic Accuracy) yielded scores from 10/25 to 20/25. Generally, the McMurray's test has relatively high specificity and low sensitivity. The studies that compared the diagnostic accuracy of the McMurray's test with that of modified versions of the test showed enhanced diagnostic accuracy for the modified tests. This review identified that the McMurray's test is of limited clinical value due to relatively low sensitivity, with modified tests (associated with the traditional McMurray's test) having higher diagnostic accuracy and thus these may be more useful clinically.

KEYWORDS: Knee, McMurray's, Meniscal, Reliability, Sensitivity, Specificity, Testing, Validity

¹Associate Professor, Health & Rehabilitation Research Centre, School of Rehabilitation and Occupation Studies, AUT University, Auckland, New Zealand

²Senior Lecturer, School of Rehabilitation and Occupation Studies, Auckland University of Technology, Auckland, New Zealand

³Head of School of Rehabilitation and Occupation Studies, Auckland University of Technology, Auckland, New Zealand

⁴School of Physiotherapy, Auckland University of Technology, Auckland, New Zealand

Address all correspondence and requests for reprints to: Wayne Hing, wayne.hing@aut.ac.nz



FIGURE 1 A) AND B): The McMurray's test: Figure 1a) the tibia rotated on the femur into full internal rotation and Figure 1b) the tibia rotated on the femur into full external rotation.

TABLE 1. Medline and CINAHL search strategy via OVID.

No.	Search history	Results
1	knee	76439
2	Menisc\$	8911
3	Mcmurray\$	140
4	Gold standard	16986
5	1 and 2	6241
6	3 and 4	5
7	3 and 5	44

Limits Human and English. \$ is the truncation character.

objective of this paper was to critically review the literature with respect to the validity and diagnostic accuracy of the traditional McMurray's test and any modifications of this test.

Method

Search Strategy

In order to make the retrieval of articles as comprehensive as possible, a generic search strategy was employed using Medline, CINAHL, and AMED databases through OVID, SPORTDiscus database through EBSCO, and SCOPUS, from 1980 to May 2008. One of the search terms used was *McMurray\$ test\$*. This generic search strategy was then combined with a subject-specific strategy (Table 1). In addition to the database searches, personal files were hand-searched by the authors for publications and relevant material. The reference lists in review articles were cross-checked

and any possibility of name/term variations was queried using MEDLINE and PUBMED.

Search Selection

All abstracts for 44 articles from Medline, 19 articles from CINAHL, 5 articles from AMED, 18 articles from SPORTDiscus, 548 articles from SCOPUS, and 6 articles from the hand search were reviewed by the authors (Figure 2). Agreement regarding which articles to read in full was determined by consensus. Studies were eligible for inclusion if they assessed measures of accuracy or validity of the McMurray's test or any modification of this test against a gold standard of either arthroscopy or magnetic resonance imaging (MRI) and were written in English. In total, 11 studies have been included in this critical review.

Methodological Analysis

Analysis of the quality of studies that evaluate the validity and accuracy of tests, such as the McMurray's test, is difficult if key information regarding the design, conduct, and analysis of the study are not reported by the authors⁹. Therefore, articles were assessed using the STARD (Standards for Reporting of Diagnostic Accuracy) checklist of methodological quality⁹, which uses established criteria for quality assessment of different research formats¹⁰. The STARD checklist contains 25 items that help to make a judgment about potential bias in the study and appraisal of the applicability of the findings. It has been used pre-

viously for the systematic assessment of the methodology of studies into diagnostic accuracy¹⁰. Three independent reviewers assessed each of the papers included in the review, and an overall STARD score of methodological quality was determined for each paper.

As previously documented in the literature¹⁰, the definition and calculation of statistical measures of concurrent criterion-validity are based on the absence or agreement between the *clinical test* and the *gold standard test*. The four possible outcomes include true positive, a false positive, a false negative, and a true negative (see Table 2). The statistical measures of sensitivity, specificity, and likelihood ratios were calculated from the information provided in the studies.

Results

Methodological Quality

The assessment results for methodological quality has been presented under the following headings: the STARD analysis, reference standard, population differences, blinding, description and interpretation of test, inter-tester reliability, diagnostic accuracy and validity, sensitivity and specificity, likelihood ratios, and McMurray's test compared to modified versions of the test.

STARD Analysis

Based on the STARD scoring of each paper, it is possible to make a qualitative assessment about the methodological quality. A consensus method was used

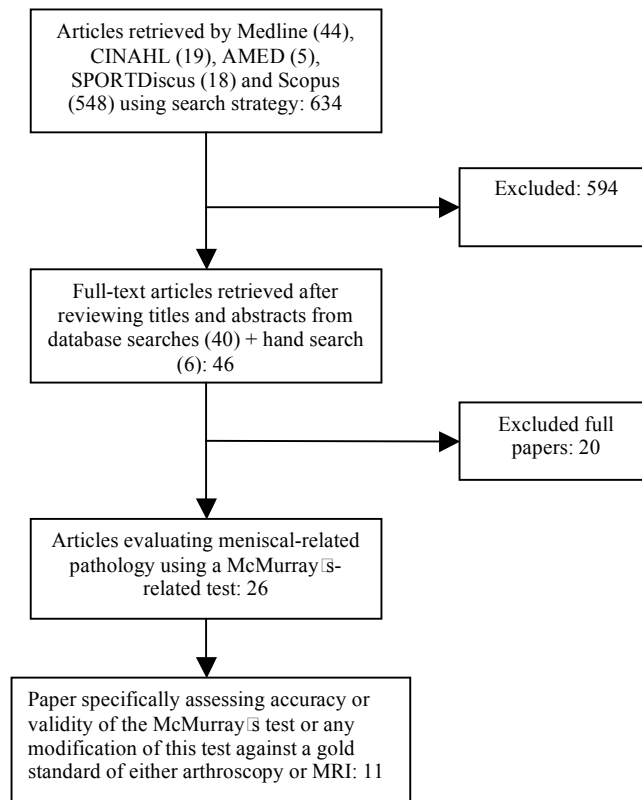


FIGURE 2. Flow diagram of literature screening process.

TABLE 2. Operational definitions of diagnostic accuracy terms used in the studies investigating validity of McMurray's test for meniscal pathology (modified from Powell & Huijbregts¹⁰).

Statistical Measure	Definition	Calculation
Sensitivity	The proportion of people who have the disease or dysfunction who test positive.	$TP / (TP + FN)$
Specificity	The proportion of people who do not have the disease or dysfunction who test negative.	$TN / (FP + TN)$
Positive Predictive Value	The proportion of people who test positive and who have the disease or dysfunction.	$TP / (TP + FP)$
Negative Predictive Value	The proportion of people who test negative and who do not have the disease or dysfunction.	$TN / (FN + TN)$
Positive Likelihood Ratio	How likely a positive test result is in people who have the disease or dysfunction as compared to how likely it is in those who do not have the disease or dysfunction.	$Sensitivity / (1 - Specificity)$
Negative Likelihood Ratio	How likely a negative test result is in people who have the disease or dysfunction as compared to how likely it is in those who do not have the disease or dysfunction.	$(1 - Sensitivity) / Specificity$
Accuracy	The proportion of true results (both <i>true positives</i> and <i>true negatives</i>) in the population. An accuracy of 100% means that the test identifies all sick and well people correctly.	$= (Number\ of\ TP + Number\ of\ TN) / (Numbers\ of\ TP + FP + FN)$

TP=true positive, FP=false positive, FN=false negative, TN=true negative

to discuss and resolve discrepancies between the markings of each paper between the three reviewers. The agreed quality for each paper is included in Table 3.

Reference Standard

Studies investigating the validity of diagnostic tests such as the McMurray's compare the findings of that test with a reference (gold) standard that has demonstrated validity¹¹. Both arthroscopy and MRI have been used as a gold standard measure for detection of meniscal injuries in knees. Arthroscopy has demonstrated an accuracy between 93% and 96%¹². There is conflicting evidence in the literature over the accuracy of MRI. A recent study by Winters and Tregonning¹³ showed a diagnostic accuracy for MRI to be 90% for the medial meniscus and 82% for the lateral meniscus. The sensitivity was 87% for the medial meniscus but only 46% for the lateral meniscus¹³. However, other studies have shown MRI to be no more accurate than clinical examination for the diagnosis of meniscal tears^{14,15}. Of the 11 studies identified in this review, nine used arthroscopy as the reference standard, one used MRI, and the remaining study used both MRI and arthroscopy (Table 4).

Population Differences

The external validity of a study is largely dependent on the study population. If a study evaluates a test in a very specific group of patients, its findings can only be applied to that same type of cohort. In testing the accuracy of a clinical test like the McMurray's test, ideally the study participants should consist of individuals who would be likely to undergo the test in clinical practice and who have a reasonable chance of having the condition¹⁶. Further, subjects who are positive on the reference standard should reflect a continuum of severity, whereas those who are negative should have conditions commonly confused with meniscal tears¹⁷. Selection bias may occur when study subjects are not representative of the population on whom the test is typically applied in practice and

can affect the results of a study¹¹. Thus, to avoid selection bias, it is important that a study include consecutive patients with pathologies that could be commonly confused with a meniscal tear and should not include patients without symptoms. The inclusion of patients with multiple pathologies is likely to lessen the diagnostic accuracy of a test; however, this would reflect actual clinical practice^{6,18}.

Six of the studies within this review included consecutive patients (Table 4). Anderson and Lipscomb⁵ used consecutive patients who were suspected of having a meniscal tear; however, these authors excluded subjects who had associated ligamentous injuries (as demonstrated by arthroscopy) from the statistical analysis. Consequently, it is likely that the accuracy of meniscal testing demonstrated by this study is artificially high compared to studies with a wider inclusion criteria.

Similarly, Corea et al⁴ included consecutive patients who were *clinically diagnosed as having torn menisci* based on a number of signs and symptoms including locking, a positive McMurray's test, painful clicks, and giving way. However, this provisional diagnosis was also based on other symptoms that one might consider could be associated with pathologies other than meniscal tears, e.g., pain, recurrent effusion, muscle wasting, and instability. These authors excluded subjects with clinical or radiographic evidence of arthritis or fracture that would increase the accuracy of testing but decrease the generalizability of the findings.

Evans et al²³ used consecutive patients on a waiting list for arthroscopy for a variety of conditions including, but not limited to, suspected meniscal tears. This was a purposeful strategy designed to enhance their ability to determine the true sensitivity and specificity of the McMurray's test in a population that reflects the symptomatic knee cohort that presents clinically. Fowler and Lubliner²² had a similarly broad population in that they included consecutive patients who warranted arthroscopic examination for any reason. However, they only included patients who had had symptoms for at least one year, making extrapolation of

their findings to the acute population challenging.

Akseki et al³ included consecutive patients with symptoms related to intra-articular knee pathology although how this was determined was not described. This study evaluated not only the McMurray's test but also a new test (Ege's test) for meniscal pathology that is performed in a weight-bearing position. Because they were investigating this weight-bearing test as well, the authors excluded any patients who presented within six weeks of trauma and those unable to bear weight or unable to squat. Once again, this affects the generalizability of the findings.

The remaining studies do not clearly state if their subjects were consecutive. Three of these studies had fairly broad inclusion criteria that better reflect the population seen in clinical practice with two including subjects with suspected meniscal or ligamentous pathology^{6,19}; the study by Sae-jung et al²⁴ included any patients identified as needing arthroscopy. The final two studies^{20,21} limited their study population to patients suspected of meniscal injury.

Blinding

Review bias may result when the findings of the reference standard test are known by the clinicians performing the diagnostic test. Knowledge of the diagnosis could influence the interpretation of the findings of the diagnostic test leading to an overstated diagnostic accuracy³. Blinding of the clinicians from the results of the diagnostic test was either not mentioned or not performed in all of the studies in this review except for the study by Karachalios et al²¹. Although these authors mentioned that the examiners were blinded to the results of the MRI, they did not make it clear if the examiners knew that there were a similar number of "normals" and symptomatic subjects included in the study or if they knew which group each individual subject belonged to. Although blinding was not mentioned in respect to the other studies, the majority required the clinical examination to be performed prior to the diagnostic arthroscopy, suggesting that the examiner

TABLE 3. STARD score of the studies retrieved.

Study	Noble 1980	Anderson 1986	Fowler 1989	Boeree 1991	Evans 1993	Corea 1994	Manzotti 1997	Kurosaka 1999	Akseki 2004	Karachailios 2005	Sae-Jung 2007
Identifies article as a study of diagnostic accuracy	1	1	1	1	1	1	1	1	1	1	1
States research questions or aims	0	1	1	0	1	0	1	1	1	1	1
Describes study population (inclusion criteria, exclusion criteria, settings, locations)	0	1	1	1	1	1	1	1	1	1	1
Describes participant recruitment	1	1	1	1	1	1	1	1	1	1	1
Describes participant sampling	1	1	1	1	1	1	1	1	1	1	1
Describes data collection (prospective or retrospective)	1	1	1	1	1	1	1	1	1	1	1
Describes reference standard and rationale	0	1	1	1	1	1	1	1	1	1	1
Describes technical specifications of material and methods involved	0	1	1	0	1	0	1	1	1	1	1
Describes definition and rationale of units, cut-off points, or categories of results of tests	0	1	1	0	1	0	1	1	1	1	1
Describes number, training, and expertise of raters	0	0	0	0	1	0	0	1	0	1	0
Were the raters blinded to the results of the other test?											
Describes clinical information available to raters	1	1	1	1	1	1	1	1	1	1	1
Describes statistical methods for comparing diagnostic accuracy and expressing uncertainty	0	0	0	1	1	1	1	1	1	1	1
Describes methods for calculating test reproducibility; if done	0	0	0	1	1	1	0	1	1	1	1
Reports when study was done with start and end dates for recruitment	0	0	0	1	0	1	1	0	0	0	1
Reports clinical and demographic characteristics of subjects	1	1	1	1	0	1	1	1	1	1	1

(continued)

Study	Noble 1980	Anderson 1986	Fowler 1989	Boeree 1991	Evans 1993	Corea 1994	Manzotti 1997	Kurosaka 1999	Akseki 2004	Karachailios 2005	Sae-Jung 2007
Reports how many subjects satisfying inclusion criteria did not undergo the tests; describes why these subjects were not tested	1	1	0	0	0	1	0	0	1	0	0
Reports time interval between researched and reference test and any treatment provided in between tests	0	0	1	0	0	0	0	0	0	0	0
Reports disease severity in subjects with target condition and other diagnoses in subjects without target condition	1	1	1	1	1	1	0	1	1	1	0
Reports cross-tabulation of researched and reference test	1	1	1	1	1	1	1	1	1	1	1
Reports adverse effects from researched and reference test	0	0	0	0	0	0	0	0	0	1	0
Reports estimates of diagnostic accuracy and measures of statistical uncertainty	0	1	1	1	1	1	1	1	1	1	1
Reports how indeterminate test results, missing responses, and outliers of researched test were handled	0	0	0	0	0	0	0	0	0	0	0
Reports estimates of variability between raters, centers, or subject subgroups; if done	0	0	0	0	1	0	0	0	0	1	0
Reports estimates of test reproducibility; if done	0	0	0	0	0	0	0	0	0	0	0
Discusses clinical applicability of study findings	1	1	1	1	1	1	1	1	1	1	1
Total Score	10	16	16	15	18	16	16	18	18	20	17

TABLE 4. Methodology and description of the 11 studies investigating validity and clinical accuracy of McMurray's test for meniscal pathology.

Study	No. of Subjects	Patient Population	Description of McMurray's Test	Objective signs of McMurray's Test	No. of Testers	Blinding Test / Arthroscopy	Reference Standard
Akseki et al ³	150	Consecutive patients. Symptoms related to an intra-articular knee pathology. Acute patients (< 6 weeks) excluded.	Described a modified version (Egés test) but no description of McMurray's.	Pain and/or click	n/m	n/m	Arthroscopy
Anderson & Lipscomb ⁵	100	Consecutive patients suspected of having meniscal tears presenting for arthroscopy: acute and chronic (ligament injuries excluded).	Described a modified version (Medial-Lateral Grind test) but no description of McMurray's.	n/m	1	n/m	Arthroscopy
Boeree & Ackroyd ¹⁹	203	Referred from GP/A&E with suspected cruciate ligament or meniscal pathology.	n/m	n/m	2 (multiple not clear)	n/m	MRI
Corea et al ⁴	93	Consecutive patients clinically diagnosed as having torn menisci (based on symptoms of pain, locking, painful clicks, recurrent effusions, giving way or signs of extension block, wasting, or instability) Patients with evidence of fracture or arthritis, a previous history of surgery, or with an acute locked knee or haemarthrosis were excluded.	Original description	Thud or click	1	n/m	Arthroscopy/ Arthroscopy
Evans et al ²³	104	Consecutive patients awaiting elective arthroscopy for suspected meniscal or other conditions based on history and physical examination. No mention acute/chronic.	Original description	Thud/sensation and/or pain	2	n/m	Arthroscopy
Fowler & Lubliner ²²	161	Consecutive patients with knee pain of at least one year's duration that warranted arthroscopic investigation.	Original description	Click/thud	2	n/m	Arthroscopy

Karachalios et al ²¹	213	Patients suspected of having a meniscal tear on the basis of history and mechanism of injury excluding those with multiple injuries, history of knee surgery, early clinical and radiographic signs of osteoarthritis, articular cartilage injuries, neurological and musculoskeletal degenerative disorders, disorders of the synovium, acute injuries (less than 4 weeks post-trauma), and any abnormal findings on conventional radiographs.	Modified McMurray's to include valgus/varus stress. Also described a weight-bearing modification of McMurray's (Thessaly test)	n/m for McMurray's, but joint line discomfort and possibly a sensation of locking or catching for Thessaly test	4	Yes	MRI and arthroscopy
Kurosaka et al ⁶	156	Patients who underwent arthroscopy to assess suspected meniscal or meniscal together with ACL injuries. All had persistent symptoms at least 8 weeks post-injury. Acute injuries excluded.	Original description	Click/thud	2	No blinding	Arthroscopy
Manzotti et al ²⁰	130	Patients diagnosed with meniscal lesions (based on symptoms including pain, recurrent edema, giving way, joint clicks, or block to movement) having arthroscopic surgery. Excluding any with past history of trauma and any with associated fractures, serious arthrosis, previous history of knee surgery or discoid meniscus identified arthroscopically.	Original description	Painful (often) click felt by examiner	4	n/m	Arthroscopy
Noble & Erat ²⁵	200	Consecutive patients scheduled for meniscectomy; acute and chronic.	n/m	n/m	1	n/m	Arthroscopy
Sae-Jung et al ²⁴	68	Patients identified as needing arthroscopy excluding those with intra-articular fracture, neurological or degenerative disorders.	Original description. Also described a modified version (the KCU compression-rotation test)	Pain or a clicking sound	n/m	n/m	Arthroscopy

n/m = not mentioned

would indeed be blinded to the results of the diagnostic test. However, only Kurosaka et al⁶ and Evans et al²³ made it clear that the examiners were not given any details about the subject's history so that they would not be influenced by this information. One study⁵ performed the test after the arthroscopy and did not state if the examiner was blinded to these results.

Description/Interpretation of Test

The description of a test within a study should be sufficient to enable replication of the test by practitioners and subsequent researchers. The description should include the exact details of the test's application and the criteria used to determine positive and negative results¹¹. Failure to do this makes it difficult to determine if the findings of the study can be compared to other studies that have evaluated the *same* test. Obviously, if the test is performed differently and/or the interpretation of a positive test is not the same, the demonstrated accuracy of the test cannot be compared.

Of the studies evaluated in this review, six used the original description of the McMurray's test^{4,6,20,22-24}. Four authors stated that they used the McMurray's test but did not describe the actual testing procedure^{3,5,19,25}. Karachalios et al²¹ incorrectly added valgus or varus stress as a component of the McMurray's. Five studies compared modified versions of the test to McMurray's^{3,5,6,21,24} (Table 3).

There are also discrepancies in the studies as to what constitutes a positive McMurray's test. Under the original description of the test, a *thud* or a *click* felt by the examiner (and sometimes heard) while performing the test was considered positive (McMurray as cited in Corea et al⁴). Other signs that have been used to denote a positive test include the production of pain, a *clunk*, or a *pop*.

Three of the studies in this review considered a positive test to be the reproduction of a palpable *thud* or *click*^{4,6,22} (Table 4). One study used a palpable *thud* and/or *pain*²³, and two studies used a palpable *click* and/or *pain*^{3,20}. Sae-Jung et al²⁴ considered pain or a clicking sound to be a positive test. The remain-

ing four studies failed to mention what denoted a positive test (Table 4). This lack of consensus in the literature highlights the risk that the criteria indicating a positive test can influence the test outcome, irrespective of whether the test was performed in the same manner on the same patient.

Intertester Reliability

The majority of studies did not report intertester or intratester reliability of the McMurray's test. Although six studies used multiple testers, these did not provide statistics for reliability^{6,19-23}. Three studies used only one tester^{4,5,25}, and two studies did not mention how many examiners were used^{3,24}. Evans et al²³ compared a senior examiner with over 10 years experience to a medical student who had recently been taught the technique whereas Karachalios et al²¹ compared two experienced orthopaedic surgeons with two inexperienced residents. Evans et al²³ demonstrated a low level of agreement between the two examiners with intertester agreements ranging from *poor* for reproduction of a medial sensation (Kappa= -0.10) to *fair* (K = +0.38) for lateral pain. They commented that the lack of intertester agreement may have been due to differences in the amount of force produced.

Evans et al²³ concluded that examiner experience had little effect on the accuracy of the diagnosis; however, they noted that the student examiner demonstrated a significant association (p= 0.002) between the diagnosis of a medial meniscus tear and reproduction of a medial thud, while the experienced examiner demonstrated a significant association between this diagnosis and the reproduction of pain (p= 0.008) or a medial "sensation" (p= 0.001). Other studies^{3,5,19} commented that greater clinical experience may affect the results of the test but they did not provide any statistical evidence to support this assertion.

These findings are contrasted by those of Karachalios et al²¹, who reported a 95% agreement for both intra- and intertester reliability for all of the clinical tests they employed. However, these authors stated that they determined these findings in a study of 20

subjects prior to the main study and they did not provide any details of how this pilot study was performed or analyzed.

Diagnostic Accuracy and Validity

Measures of efficacy include accuracy, sensitivity, and specificity. Accuracy is the percentage of subjects who are correctly identified as either having or not having a meniscal tear. The accuracy measure has limited usefulness in that it does not distinguish between the diagnostic value of positive and negative results¹¹. To some degree, this is achieved by sensitivity and specificity, which provide useful information for interpreting the results of diagnostic tests.

Sensitivity and Specificity

Sensitivity can be defined as the proportion of patients with the condition who have a positive test result and represents the ability of the test to recognize the condition when present¹¹. Specificity is the proportion of patients without the condition who have a negative test result and indicates the ability to use a test to recognize when the condition is absent¹¹. High sensitivity indicates that a test can be used for excluding a condition when it is negative, but it does not address the value of a positive test. High specificity indicates that a test can be used for including a condition when it is positive²⁶. Sensitivity and specificity rely on a single threshold for classifying a test result as positive or negative. Changing the threshold to increase sensitivity decreases specificity and vice versa. This trade-off between sensitivity and specificity makes it important that they be considered jointly²⁷. This means that tests rarely have both high sensitivity and specificity.

As is true of all statistics, sensitivity and specificity values are taken from a sample and represent an estimate of the true value that could be found in the population. The confidence interval (CI) attests to the precision of this estimate¹¹. A 95% CI is the most commonly used and indicates a range of values within which the population value would lie with 95% certainty. If the CI is wide and contains values that are not

clinically important, the usefulness of the measure may be questionable¹¹.

The sensitivity and specificity of the McMurray's test reported in the studies identified in this review vary widely (Table 5). Sensitivity figures vary from 16%–88%, while specificity figures vary from 20%–98% (Table 5). In general, sensitivity figures are much lower than specificity and the CI limits are wider. Sensitivity figures were higher than specificity for three studies^{5,20,25} (Table 5). The low sensitivity figures would indicate that in general, a negative test result is not reliable in ruling out meniscal pathology and a torn meniscus would likely be missed if the McMurray's test was the sole determinant of pathology. Higher specificity figures denote that in general when the McMurray's test is positive, it is fairly reliable for ruling in meniscal pathology.

Fowler and Lubliner²² attributed their low sensitivity results (compared to previous studies)^{5,25} to population differences between the studies (Table 5). This was also discussed by Evans et al²³, who attributed their low sensitivity rates to wide patient entry criteria including differing pathologies (Table 4).

A recent study by Akseki et al³ reported high combined sensitivity and specificity figures (63% and 83%, respectively) and relatively narrow confidence intervals (Table 5). These authors suggested that this increase in sensitivity and specificity compared to previous studies was due to their broader definition of a positive test, i.e., reproduction of a click or pain³; however, this does not explain the similar findings of Corea et al⁴ in which only a click was indicative of a positive test.

Some of the studies did not separate the data for medial from that of lateral meniscal testing^{5,6,22,25}. However, of those that have made this distinction, there is some consensus that the McMurray's test has higher sensitivity with respect to medial meniscal tears and higher specificity with lateral meniscal tears^{3,4,19,20,24}.

Likelihood Ratios

Although sensitivity and specificity values provide useful information, they work against the direction of clinical

testing¹¹. Clinically, we do not know whether a patient has the condition before the diagnostic test (arthroscopy or MRI) is performed. Sensitivity and specificity values infer the probability of a correct test, given the result of the reference standard¹¹. They also fail to take into account pre-test probability. Useful tests should produce large shifts in probability once the result of the test is known. Sensitivity and specificity values fail to do this¹¹. The best statistics for summarizing usefulness of a diagnostic test appear to be likelihood ratios (LR)¹⁷. Likelihood ratios overcome some of the problems involved with sensitivity and specificity values by summarizing the information contained in these values in a manner that can be used to quantify shifts in probability once the meniscal test results are known²⁸.

An LR+ indicates the degree of certainty that a patient with a positive test actually has the suspected condition while an LR- indicates the degree of certainty that a patient with a negative test does not have the suspected condition²⁷. An LR of 1 indicates that the test result does nothing to change the likelihood that the patient either does or does not have the condition, whereas the higher the LR+, the more certain you can be that a positive test indicates the person has the disorder. The lower the LR-, the more certain you can be that a negative test indicates the person does not have the disorder¹¹ (Table 6). An example of this would be as follows: If the McMurray's test had a LR+ of 9.2 for a particular study, a positive McMurray's test is 9.2 times more likely to occur in patients with a meniscal tear than in those without one²⁹.

Table 5 shows the LR+ and LR- for the 11 studies included within this review with 95% CIs. The wide range of positive likelihood ratios (0.82–8.86) make it difficult to draw any conclusions about the actual magnitude of this ratio. Four studies demonstrated that a positive test alters the probability to only a small, rarely important degree^{5,6,25}, suggesting uncertainty that a positive test will indicate meniscal pathology (Table 5). Studies by Boeree and Ackroyd¹⁹, Akseki et al³, and Karachalios et al²¹ demonstrated small but sometimes impor-

tant shifts in probability. Unfortunately, it is not possible to accurately determine the precision of reliability of the Boeree and Ackroyd¹⁹ study as CIs could not be calculated. Of the four studies that demonstrated the highest shifts in probability, only Corea et al⁴ and Akseki et al³ contained calculable CIs, which were relatively narrow (Table 5).

With regard to negative likelihood ratios, all but three of the studies demonstrated only a small alteration in probability that a subject with a negative McMurray's test will not have a meniscal tear (Table 5). In one of these studies, the CIs are extremely wide⁵. However, in general, the CI limits are relatively narrow over all. The studies by Akseki et al³, Corea et al⁴, and Manzotti et al²⁰ revealed negative likelihood ratios that are slightly lower than the other studies. These represent small but sometimes important shifts in probability and the stronger methodology of these studies is reflected by the relatively narrow CIs (Table 5).

McMurray's Test Compared to Modified Versions

Some studies have attempted to compare the diagnostic value of the McMurray's test to that of modified tests. These studies have hypothesized that by incorporating aspects of varus/valgus stress and/or axial loading into the original McMurray's test, there is an increase in diagnostic value^{3,5,6}.

Anderson and Lipscomb⁵ compared the McMurray's test to a test termed the Medial-Lateral Grind test that included a varus/valgus component not included in the original McMurray's test. The Medial-Lateral Grind test had a higher LR+ (Table 7) when compared to the McMurray's test; however, its CIs were extremely wide, bringing into question the precision of this estimate of reliability (Table 7). These authors also demonstrated that the Medial-Lateral Grind test had smaller (better) LR- compared to the McMurray's test although the change in probability was still only small and should be considered rarely important (Table 7).

Kurosaka et al⁶ took the modification of the Medial-Lateral Grind test

TABLE 5. Sensitivity, specificity, and likelihood ratios (LR) of the McMurray's test with confidence intervals (CI).

Study	Sensitivity (%) CI (%)			Specificity (%) CI (%)			LR+	LR-
	Medial & Lateral combined = meniscal tear	Medial meniscus	Lateral meniscus	Medial & Lateral combined = meniscal tear	Medial meniscus	Lateral meniscus	(CI 95%)	(CI 95%)
Akseki et al ³	63 (55-71)	67 (59-75)	53 (45-61)	83 (77-89)	69 (62-76)	88 (83-93)	3.71 (3.19-4.13)	0.45 (0.39-0.52)
Anderson & Lipscomb ⁵	58* (48-68)			29* (20-38)			0.82 (0.5-1.3)**	1.45 (0.4-4.9)**
Boeree & Ackroyd ¹⁹	27* (21-33)	29.3 (23-36)	25 (19-31)	89* (85-93)	87.3 (83-92)	89.8 (86-94)	2.31*	0.81*
Corea et al ⁴	58.5 (48-69)	65 (55-74)	52 (41-62)	93.4 (88-98)	93 (88-98)	94 (88-99)	8.86 (7.17-10.91)	0.44 (0.36-0.54)
Evans et al ²³	33* (24-42)	16 (9-23)	50 (40-60)	96* (92-100)	98 (95-100)	94 (89-99)	8.33*	0.70*
Fowler & Lubliner ²²	29* (22-36)			96* (93-99)			7.25*	0.74*
Karachalios et al ²¹		48	65		94	86	8.00 4.64	0.553 0.40
Kurosaka et al ⁶	37 (30-44)			77 (70-84)			1.61*0.82*	
Manzotti et al ²⁰		88	79		50	20	1.76 0.98	0.24 1.05
Noble & Erat ²⁵	63 (56-70)			57 (50-64)			1.50 (1.1-2.1)**	0.60 (0.5-0.9)**
Sae-Jung et al ²⁴	70.6 (55-82)	70	68.2	82.49 (55-95)	60.7	47.8	4.00 (1.4-11.3)	0.358 (0.22-0.56)

When confidence intervals are not present, the CIs were incalculable due to absence of raw data.

** No raw data available to calculate CIs values taken from Solomon et al³¹.

*CIs calculated using formulas from Sackett et al³².

TABLE 6. A guide to the interpretation of likelihood ratio (LR) values.

LR+	LR-	Interpretation
>10	<0.1	Generate large and often conclusive shifts in probability
5-10	0.1-0.2	Generate moderate shifts in probability
2-5	0.2-0.5	Generate small but sometimes important shifts in probability
1-2	0.5-1	Alter probability to a small and rarely important degree

Modified from Jaeschke et al¹⁷.

further by comparing the McMurray's test to a pivot shift test that not only had a component of varus/valgus stress but also included a component of axial loading. These authors considered the overall accuracy of the axially loaded pivot shift test to be higher than that of the McMurray's test (Table 7). Confidence

intervals could not be calculated³² from the data provided by these authors making it difficult to assess the accuracy of results.

Akseki et al³ compared the McMurray's test with a weight-bearing version of the McMurray's test that incorporated axial compression and varus/valgus

stress, with the patient squatting down in internal and then external rotation (Ege's test). The modified weight-bearing test showed a higher LR+ and a lower LR- than the McMurray's test (Table 7). These results have been supported by Karachalios et al²¹, who compared another weight-bearing modification (the Thessaly test) of the McMurray's with the original test. These authors demonstrated significantly larger (better) positive likelihood ratios and significantly smaller (better) negative likelihood ratios than the McMurray's.

The final study by Sae-Jung et al²⁴ compared a modified version to McMurray's added axial compression, similar to that applied by Kurosaka et al⁶ but

TABLE 7. Comparison of likelihood ratio's for McMurray's test with modified tests.

Study	LR	McMurray's		Modified			
		General	Medial meniscus	Lateral meniscus	General	Medial meniscus	Lateral meniscus
Akseki et al ³	LR+	3.71 (3.19–4.13)			5.15 (4.48–5.93)		
	LR–	0.45 (0.39–0.52)			0.38 (0.33–0.44)		
Anderson & Lipscomb ⁵	LR+	0.82 (0.5–1.3)**			4.8 (0.8–30.0)		
	LR–	1.45 (0.4–4.9)**			0.4 (0.2–0.6)		
Karachalios et al ²¹	LR+		8.0	4.64		26.8 (14–51)	23.0 (13–37)
	LR–		0.55	0.40		0.11 (0.06–0.18)	0.08 (0.02–0.2)
Kurosaka et al ⁶	LR+	1.61*			4.18		
	LR–	0.82*			0.35		
Sae-Jung et al ²⁴	LR+		1.63	1.64		1.77	1.95
	LR–		0.69	0		0.33	0

* CIs incalculable due to absence of raw data.

** No raw data available to calculate CIs figures taken from Solomon et al 2001³¹.

without added valgus or varus stress. These authors demonstrated marginally better LR+ but most interestingly, reported that their modified test (the KKU test) was 100% sensitive for lateral meniscal tears indicating that the test can be used for excluding a condition when it is negative.

While it is difficult to compare results across studies due to the differences in the tests being used, the results of this review appear to show that the modified tests have higher diagnostic value than the McMurray's test.

Discussion

On the basis of the results of the studies in this review, it seems that intertester reliability using the McMurray's test is low. This is not surprising given the complicated nature of the technique and the difficulty in controlling the amount and direction of forces across testers. It is important to take this into consideration when analyzing test results of studies that have used more than one examiner. While some studies have stated that greater clinical experience aids correct diagnosis^{3,5,19}, the only current statistical evidence in this regard shows no difference between an experienced and inexperienced tester²³.

Similarly, sensitivity figures ranged from 27% to 70% across the reviewed papers, generally indicating that a torn meniscus is likely to be missed in many patients; however, specificity figures (29–96%) indicating that false positive tests are relatively low and that a positive test makes it likely that the patient actually does have a torn meniscus. Results also indicate that testing for medial meniscal pathology is more sensitive than testing for lateral; however, tests for lateral meniscal pathology are more specific than tests for medial pathology^{3,4,19}. In contrast, the paper by Sae-Jung et al²⁴ found sensitivity for medial and lateral menisci of 70% and 68%, respectively, and specificity values for medial and lateral menisci of 60.7% and 47.8%, respectively. Unlike the medial meniscus, which is attached to the medial ligament, the lateral meniscus is not attached to the lateral ligament. Mariani et al³⁰ have suggested that the differences in anatomical attachments of the two menisci contribute to these variations in sensitivity and specificity of diagnostic tests³⁰.

Positive likelihood ratios presented in the studies reviewed generally indicated small to moderate shifts in probability (0.82–8.86) in that a positive test will indicate true meniscal pathology

although the studies with the highest methodological quality demonstrated likelihood ratios considered to indicate moderate improvements in the probability that this will be the case^{3,4}. Relatively narrow confidence intervals also attest to the reliability of these two studies^{3,4} (Table 5).

The differences in study populations are likely to have contributed to the wide variability of results across studies. Those that do not include consecutive patients and those that exclude different pathologies may have biased results. There is conflicting evidence with respect to the effect of the presence of an associated anterior cruciate ligament (ACL) deficiency. Kurosaka et al⁶ stated that diagnostic accuracy is lessened in patients with multiple pathologies, whereas Akseki et al³ found that there was no reduction in diagnostic accuracy with an associated tear of the ACL. The inclusion of patients with different pathologies would make the results of studies more generalizable to the clinical setting.

The varying definitions of a positive McMurray's test are also likely to have contributed to the variability of the results demonstrated by the studies reviewed. It seems logical that those studies that include both pain and a click

should have higher diagnostic value as compared to studies that just use one sign or the other. This is true in the case of the study by Akseki et al³ but not for the study by Evans et al²³ (Tables 4 and 5).

Differences in the type of tear have been suggested as influencing the result of clinical tests; however, no detailed investigation of this issue exists in the current literature³. McMurray clearly indicated that the test that bears his name is only relevant for tears in the posterior portion of the cartilage (McMurray, 1942, cited in Corea et al⁴).

A recent literature review on composite testing of the diagnostic tests for the meniscus reported reasonable sensitivity and specificity when the findings of a number of tests are combined³¹. This, along with the conclusions discussed above, suggests that the McMurray's test should be used as one of a combination of tests in the clinical setting^{3,22,23}.

Three studies in this review compared the McMurray's test to modified versions that incorporated the added components of varus/valgus stress and axial compression. Each of these studies demonstrated improved diagnostic accuracy of these modified tests compared to the original McMurray's; however, they concluded that the modified tests should be used as well, as rather than as an alternative to other diagnostic tests^{3,5,6}. One problem with these modified tests is that they appear to have all been evaluated by the creators of the tests, which to some degree challenges the validity of the research. These comments are also supported by the findings of a recent meta-analysis carried out by Hegedus et al⁷ and Meserve et al⁸. These authors also observed that the studies on these new tests have only been subjected to scientific scrutiny on one occasion and further research is required on these tests.

Limitations

Limitations of this review relate to the search strategy used. Articles may have been missed based on the omission of certain search phrases or the use of a single search phrase as used in this case. Limiting the search to English language

articles only may also have led to an omission of other relevant studies. Studies were also not examined where they clearly did not meet the search criteria.

The use of the STARD tool is also a limitation. This is a relatively new tool and has not been subjected to an analysis of its reliability at this time; however, the tool does provide a consistent framework on which to base the analysis of diagnostic studies. The preliminary nature of this tool also means that a more narrative review of the validity and accuracy of the tests has been presented.

Research and Clinical Implications

Future research should concentrate on building a strong methodological base incorporating large samples of consecutive patients with commonly confused pathologies. Further, the description of the test itself should be well explained, and improving intertester reliability in the future would increase the validity of the studies. Finally, further independent research needs to compare the McMurray's test with modified tests to confirm the apparent superiority of these tests over the McMurray's test.

Bearing these findings in mind, the following recommendations can be made for the clinician:

1. Be aware of the validity issues surrounding this test.
2. Consider reproduction of pain during the test as a positive test, not just the reproduction of a *click* or *thud*.
3. Consider the findings of this test in conjunction with those of other tests to enhance the likelihood of a correct diagnosis such as joint line tenderness.
4. Consider the use of modifications of the test for improved validity.

Conclusion

This review has demonstrated that the intertester reliability and sensitivity of the McMurray's test is relatively low; however, it has also highlighted that it can be a relatively specific test, especially with respect to the lateral meniscus. The

review suggests that modifications of the interpretation of a positive test to include reproduction of pain either as well as or on its own may enhance the validity of the test. The review also highlights the idea that modified versions of the test seem to be more valid than the original version.

REFERENCES

1. Malanga GA, Andrus S, Nadler SF, McLean J. Physical examination of the knee: A review of the original test description and scientific validity of common orthopedic tests. *Arch Phys Med Rehabil* 2003;84:592-603.
2. ACC. *The Management of Soft Tissue Knee Injuries: Internal Derangements*. Wellington, NZ: A.C. Corporation, 2003.
3. Akseki D, Ozcan O, Boya H, Pinar H. A new weight-bearing meniscal test and a comparison with McMurray's test and joint line tenderness. *Arthroscopy* 2004;20:951-958.
4. Corea JR, Moussa M, Al Othman A. McMurray's test tested. *Knee Surg Sports Traumatol Arthroscop* 1994;2:70-72.
5. Anderson AF, Lipscomb AB. Clinical diagnosis of meniscal tears: Description of a new manipulative test. *Am J Sports Med* 1986;14:291-293.
6. Kurosaka M, Yagi M, Yoshiya S, Muratsu H, Mizuno K. Efficacy of the axially loaded pivot shift test for the diagnosis of a meniscal tear. *Int Orthop* 1999;23:271-274.
7. Hegedus EJ, Cook C, Hasselblad V, Goode A, McCrory DC. Physical examination tests for assessing a torn meniscus in the knee: A systematic review with meta-analysis. *J Orthop Sports Phys Ther* 2007;37:541-550.
8. Meserve BB, Cleland JA, Boucher CT. A meta-analysis examining clinical test utilities for assessing meniscal injury. *Clin Rehabil* 2008;22:143-161.
9. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *BMJ* 2003;326:41-44.
10. Powell JW, Huijbregts PA. Concurrent criterion-related validity of acromioclavicular joint physical examination tests: A systematic review. *J Man Manip Ther* 2006;14:E19-E29.
11. Fritz JM, Wainner RS. Examining diagnostic tests: An evidence-based perspective. *Phys Ther* 2001;81:1546-1564.
12. Muellner T, Weinstabl R, Schabus R, Vecsei V, Kainberger F. The diagnosis of meniscal

- tears in athletes: A comparison of clinical and magnetic resonance imaging investigations. *Am J Sports Med* 1997;25:7-12.
13. Winters K, Tregonning R. Reliability of magnetic resonance imaging of the traumatic knee as determined by arthroscopy. *NZ Med J* 2005;118:1-8.
 14. Miller GK. A prospective study comparing the accuracy of the clinical diagnosis of meniscal tears with magnetic resonance imaging and its effect on clinical outcome. *Arthroscopy* 1996;12:406-413.
 15. Rose NE, Gold SM. A comparison of accuracy between clinical examination and magnetic resonance imaging in the diagnosis of meniscal and anterior cruciate ligament tears. *Arthroscopy* 1996;12:398-405.
 16. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: Getting better but still not good. *JAMA* 1995;274:645-651.
 17. Jaeschke RZ, Meade MO, Guyatt GH, Keenan SP, Cook DJ. How to use diagnostic test articles in the intensive care unit: Diagnosing weanability using f/Vt. *Crit Care Med* 1997;25:1514-1521.
 18. Oberlander MA, Shalvoy RM, Hughston JC. The accuracy of the clinical knee examination documented by arthroscopy: A prospective study. *Am J Sports Med* 1993;21:773-778.
 19. Boeree NR, Ackroyd CE. Assessment of the menisci and cruciate ligaments: An audit of clinical practice. *Injury* 1991;22:291-294.
 20. Manzotti A, Baiguini P, Locatelli A, et al. Statistical evaluation of McMurray's test in the clinical diagnosis of meniscus injuries. *J Sports Traumatol Relat Res* 1997;19:83-89.
 21. Karachalios T, Hantes M, Zibis AH, Zachos V, Karantanas AH, Malizos KN. Diagnostic accuracy of a new clinical test (the Thessaly Test) for early detection of meniscal tears. *J Bone Joint Surg* 2005;87:955-962.
 22. Fowler PJ, Lubliner JA. The predictive value of five clinical signs in the evaluation of meniscal pathology. *Arthroscopy* 1989;5:84-86.
 23. Evans PJ, Bell GD, Frank CY. Prospective evaluation of the McMurray test. *Am J Sports Med* 1993;21:604-608.
 24. Sae-Jung S, Jirattanaphochai K, Benjasil T. KKU knee compression-rotation test for detection of meniscal tears: A comparative study of its diagnostic accuracy with the McMurray test. *J Med Assoc Thailand* 2007; 90:718-725.
 25. Noble J, Erat K. In defence of the meniscus: A prospective study of 200 meniscectomy patients. *J Bone and Joint Surg BR* 1980;62:7-11.
 26. Schulzer M. Diagnostic tests: A statistical review. *Muscle Nerve* 1994;17:815-819.
 27. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med* 1994;120: 667-676.
 28. Simmel DL, Samsa GP, Matchar DB. Likelihood Ratios with confidence: Sample size estimation for diagnostic test results. *J Clin Epidemiol* 1991;44:763-770.
 29. Bhandari M, Guyatt GH. How to appraise a diagnostic test. *World J Surg* 2005;29:61-66.
 30. Mariani PP, Adriani E, Maresca G, Mazzola CG. A prospective evaluation of a test for lateral meniscal tears. *Knee Surg Sports Traumatol Arthroscop* 1996;4:22-26.
 31. Solomon DH, Simel DL, Bates DW, Katz JN, Schaffer JL. Does the patient have a torn meniscus of ligament of the knee? Value of the physical examination. *JAMA* 2001;286: 1610-1620.
 32. Sackett D, Richardson S, Rosenberg W, Haynes RB. *Evidence-Based Medicine: How to Practice and Teach EBM*. Edinburgh, UK: Churchill-Livingstone, 1996.