



Special tests for assessing meniscal tears within the knee: a systematic review and meta-analysis

Benjamin E Smith,¹ Damian Thacker,² Ali Crewesmith,¹ Michelle Hall³

10.1136/ebmed-2014-110160

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/ebmed-2014-110160>).

¹Department of Physiotherapy Outpatients, London Road Community Hospital, Derby Hospitals NHS Foundation Trust, Derby, UK

²Department of Physiotherapy Outpatients, Ashfield Health Village, Kirkby-In-Ashfield, Nottingham, UK

³School of Health Sciences, Clinical Sciences Building, University of Nottingham, Nottingham, UK

Correspondence to Benjamin E Smith, Department of Physiotherapy Outpatients, London Road Community Hospital, Derby Hospitals NHS Foundation Trust, London Road, Derby DE1 2QY, UK; benjamin.smith3@nhs.net



Editor's choice
Scan to access more
free content

Abstract

Background Musculoskeletal knee pain is a large and costly problem, and meniscal tears make up a large proportion of diagnoses. 'Special tests' to diagnose torn menisci are often used in the physical examination of the knee joint. A large number of publications within the literature have investigated the diagnostic accuracy of these tests, yet despite the wealth of research their diagnostic accuracy remains unclear.

Aim To synthesise the most current literature on the diagnostic accuracy of special tests for meniscal tears of the knee in adults.

Method An electronic search of MEDLINE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), The Allied and Complementary Medicine Database (AMED) and SPORTDiscus databases was carried out from inception to December 2014. Two authors independently selected studies and independently extracted data. Methodological quality was evaluated using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS) 2 tool.

Results Nine studies were included (n=1234) and three special tests were included in the meta-analysis. The methodological quality of the included studies was generally poor. McMurray's had a sensitivity of 61% (95% CI 45% to 74%) and a specificity of 84% (95% CI 69% to 92%). Joint line tenderness had a sensitivity of 83% (95% CI 73% to 90%) and a specificity of 83% (95% CI 61% to 94%). Thessaly 20° had a sensitivity of 75% (95% CI 53% to 89%) and a specificity of 87% (95% CI 65% to 96%).

Conclusions The accuracy of the special tests to diagnose meniscal tears remains poor. However, these results should be used with caution, due to the poor quality and low numbers of included studies and high levels of heterogeneity.

Introduction

The lifetime prevalence of musculoskeletal knee pain within England is 54%,¹ with the point prevalence in working adults over 40 years of age 28%.² Within the UK, 24% of workers between the ages of 16 and 65 present with musculoskeletal knee pain lasting up to 2 years, with 12% of all workers saying they needed time off within the past 12 months due to knee pain.³ Despite its importance to clinicians and patients, there is a paucity of information on the epidemiology of meniscal tears.⁴ The incidence of specific meniscal tears within the Netherlands is 2/1000/year,⁵ and they account for 25 000 hospital admissions a year in the UK.⁶ Although the prevalence of specific meniscal injuries within the UK is unknown, the point prevalence has been recorded as 57% in symptomatic knees and 36% in asymptomatic knees in

Switzerland,⁷ and 32% in symptomatic knees and 23% in asymptomatic knees in the USA.⁸

'Special tests' have been a historical part of the physical examination during the clinical assessment of musculoskeletal knee pain,⁹ and a number of these special tests are thought to diagnose torn menisci. Apley's, McMurray's and joint line tenderness (JLT) are commonly used in practice,¹⁰ with Thessaly's being considered a new dynamic test with high diagnostic accuracy.¹¹ The diagnostic accuracy of these special clinical tests for the detection of meniscal tears has been examined quite extensively within the literature, yet still remains unclear.^{9 12–15} Previous systematic reviews have not limited the age range of included participants to adults only, with many of the studies including children within their data. In addition, there exists some confusion over the definitions of the test procedures.^{9 12–15} For example, McMurray's test was originally described with the knee being tested from full flexion to 90°,¹⁶ but its use and application now varies widely.¹⁷ Similarly, Apley's test is originally described as only applying a lateral rotation force,¹⁸ but is often described with a lateral and medial rotational force.^{11 19–21} Four of the five previous systematic reviews do not make clear their specific definitions of their test procedure, and nor do they make an attempt to analyse or categorise their included studies by their definition of the clinical special tests used.^{12–15} Hegedus *et al*⁹ did attempt to subcategorise and analyse studies by test definition. However, it is unclear how much investigative work was carried out. For example, they included Karachalios *et al*¹¹ in which clear contradiction exists for two of their special tests, since they reference two publications that describe them in different ways. Hegedus *et al*⁹ did not state in their data synthesis how this confusion was dealt with.

The last systematic review on the diagnostic accuracy of special tests for meniscal tears was conducted almost 6 years ago, with unclear results.¹⁵ Since then, the literature has been greatly added to; new standards of methodological quality by which systematic reviews are measured against have been introduced,²² and the statistical method by which meta-analyses are carried out for diagnostic accuracy studies has improved with the unification of the bivariate model.^{23 24}

The main objective of this review was to synthesise the most up-to-date literature for diagnostic accuracy studies for meniscal tears of the knee for adults specifically and, if the data allowed, pool results into a meta-analysis.

Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were utilised during the search and reporting phase of this systematic review/meta-analysis.²²

Search strategy

An electronic database search of titles and abstracts was conducted from inception to December 2014. A systematic search of the literature was conducted using MEDLINE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), The Allied and Complementary Medicine Database (AMED) and SPORTDiscus databases. The specific search strategy differed depending on the electronic database being searched at that time (see online supplementary table S1 for the MEDLINE keywords and search strategy). Relevant articles, titles and abstracts were identified and screened and the reference lists of retrieved articles were also searched for additional references. An attempt was made to identify unpublished studies by emailing all authors from retrieved studies and previous systematic reviews.

Eligibility criteria

All studies examining the accuracy of special tests in diagnosing meniscal tears of the knee in adults (16 years of age or older) were included. The study must have had at least one clinical special test, must have reported specificity and sensitivity and been written in English. Special tests included McMurray's test,¹⁶ Apley's test,¹⁸ Thessaly's test¹¹ or JLT.¹⁰ The tests must not have been carried out under anaesthetics or on cadavers, or been part of a composite examination. Clinical diagnosis by MRI or arthroscopy surgery was considered the gold standard reference test.

Study selection

One reviewer (BES) conducted the initial database searches and screened the titles and abstracts. Full copies of potential eligible papers were retrieved and independently screened by two reviewers (BES and DT). The initial percentage agreement was 95%. Using Cohen's statistical method, κ agreement was $\kappa=0.87$, which is considered near-perfect agreement.^{25–26} Any disagreements were resolved through discussion, without the need for a third reviewer who was available (AC).

Data extraction

One reviewer (BES) independently extracted data regarding study design, participant information, gold standard test used, clinical special test information, setting and outcome data.²⁷ All data were independently checked by a second reviewer (AC). A description of the examination protocols for each special test is included in table 1.

In order to complete pooling of data through a meta-analysis, the raw 2x2 data are required.²⁸ Of the included studies, five had incomplete data to allow for this.^{11–21–29–31} All five were contacted, two responded and provided the raw data,^{30–31} one responded but advised that they no longer had the data,²¹ and two failed to respond.^{11–29} Sensitivity, specificity and likelihood ratios were calculated and summarised in table 2 along with the study characteristics.

Quality assessment

The methodological quality of the included studies was assessed independently by both reviewers using the Quality Assessment Tool for Diagnostic Accuracy

Studies (QUADAS) 2 tool (BES and DT).³² Despite there being no point scoring system with QUADAS 2, authors may restrict primary analysis to studies only showing a low risk of bias; however, this is not thought to be best practice and subgroup analysis postheterogeneity investigation is considered optimal.³² Disagreements were resolved through discussion, with a third reviewer available (AC). Results were presented through graphs and tables provided by QUADAS via their website.³²

Statistical analysis

All data were analysed using the OpenMetaAnalyst software.²⁸ Heterogeneity between studies was assessed through the I^2 statistic, with this systematic review considering 25% low, 50% moderate and 75% high.³³ 2x2 Data tables were created in order to perform a meta-analysis. If data were unavailable, then studies remained within the review for qualitative analysis. As recommended by Harbord's unification of models for meta-analysis of diagnostic accuracy studies, the bivariate model was used for pooling of data with their corresponding 95% CIs.^{23–24} The bivariate model is recommended for diagnostic accuracy studies where inherent heterogeneity exists between studies, for example, in threshold effects, study populations and index testing protocols, since it uses a model similar to the random effects model used for treatment efficacy meta-analyses.²⁴

Sensitivity, also called the true positive rate, is the measure of true positives actually identified (eg, the percentage of people with a meniscal tear who are correctly diagnosed as having a tear). Specificity, also called the true negative rate, is the measure of true negatives actually identified (eg, the percentage of people who do not have a meniscal tear, who are correctly diagnosed as not having one). These two measures are combined to give likelihood ratios. The positive likelihood ratio (LR+) is a measure of how much the probability of having a tear increases in the presence of a positive test result. An LR+ of 1 indicates that the post-test probability is exactly the same as the pretest probability, and greater than 1 indicates that the probability has increased.³⁴ The higher the LR+, the greater the probability increase.³⁴ The negative likelihood ratio (LR-) is a measure of how much the probability of having a tear decreases in the absence of a positive test result.³⁵ An LR- of below 1 indicates that the post-test probability has decreased, and the smaller the LR- the greater the decrease in probability.³⁴

Where primary analysis was restricted to studies only showing a low risk of bias, the robustness of our results was tested through a sensitivity analysis.

Results

Study identification

The initial database search produced 739 citations, with 6 further studies found through reference list searches. Only one unpublished trial was identified, but unfortunately they declined to allow it to be included within this review. After duplicates were removed, 43 were appropriate for full-text review (see figure 1 for study selection process).

After full-text review, 26 studies were excluded due to participants not meeting the criteria (all due to participants not exclusively being 16 years or older);^{36–61} 9 due

Table 1 Test procedures of included studies

	McMurray's	Apley's	Joint line tenderness	Thessaly
Akseki <i>et al</i> ²⁹	No specific details given, original test referenced ¹⁶		No specific details given	
Corea <i>et al</i> ⁶⁷	No specific details given, original test referenced ¹⁶			
Eren ³⁰			Performed at 90° flexion. No definition of 'tenderness' for positive test	
Galli <i>et al</i> ³¹	Original text referred to, cited in Malanga <i>et al</i> ¹⁰		Original text referred to, cited in Malanga <i>et al</i> ¹⁰ . Does not include a definition of 'tenderness' for positive test	
Karachalios <i>et al</i> ¹¹	No specific details given, original test referenced, ¹⁶ plus 'Clinical examination of the knee' referenced ¹⁹	No specific details given, referenced to Tria in 'Clinical examination of the knee' ¹⁹	No specific details given, but Tria in 'Clinical examination of the knee' referenced ¹⁹	Details: Patient's hands held while they single leg stand and twist their body and rotate their knee internally and externally×3. In 5° and 20°. Positive=pain, locking or catching
Konan <i>et al</i> ⁵³	No specific details given, original test referenced ¹⁶		No test description, nor referenced	Description given similar to that given by Karachalios <i>et al</i> ¹¹
Manzotti <i>et al</i> ⁶⁸	Description given, in line with original test. But not detailing how far to extend the knee. Original test also referenced ¹⁶			
Mirzatoiloei <i>et al</i> ⁶⁹	No test description, nor referenced		No test description, nor referenced	No description given, but referenced to Karachalios <i>et al</i> ¹¹
Rinonapoli <i>et al</i> ²¹	Description given, in line with original test. But not detailing how far to extend the knee. Original test also referenced ¹⁶	Description given, in line with original test. But not detailing which way to rotate the knee. Original test also referenced ¹⁸		

to the study design not meeting the criteria;^{52 55 58–60 62–64} and 2 due to no outcome data being recorded.^{65 66} Some of the studies were excluded for more than one reason. Nine studies remained for full inclusion of this review.^{11 21 29–31 53 67–69}

Characteristics of included studies

There were heterogeneous populations within the included studies with regard to age, duration of symptoms and sex (table 2). Age of participants of the studies varied widely, with the mean age ranging from 19.2 to 39 years of age.^{30 53} One study did not publish age ranges, but confirmed that all were adults via email.³¹ Five studies did not specify the mean duration of symptoms,^{11 21 53 67 69} and in the remaining four studies the symptom duration ranged from 14³⁰ to 52 months.³¹ Heterogeneity also existed with regard to the description and specific manoeuvre of these special tests (table 1).

Study quality and bias

Percentage agreement between the two reviewers for the overall score using the QUADAS 2 tool was 75%, with a κ of $\kappa=0.43$, which is considered moderate to fair.^{25 26} No study received a score of high risk of bias in more than one of the four categories (figure 2).

The greatest risk of bias was with the index test and patient selection. There was a great amount of uncertainty with regard to reference testing, as most studies failed to explicitly say that this was carried out without knowledge of the index test results. There were often inappropriate exclusion,^{21 29} and inappropriate patient selection.³⁰ Flow and timing issues were associated with poor documentation, as most failed to specify the length

of time between index and reference test, or failed to confirm that all patients received the reference test and were included within the analysis. With all but one study,¹¹ the greatest source of bias was with regard to verification bias, as all participants were referred for secondary care with knee symptoms and suspected meniscal tears. All nine studies were included within quantitative data synthesis.

Data synthesis

Nine studies, with 1234 participants, were included within this review. Apley's test¹⁸ was investigated by two studies,^{16 18} both of which failed to supply the raw 2×2 data. Thessaly at 5°¹¹ was investigated by two studies,^{16 31} but only Konan *et al*⁵³ supplied the raw 2×2 data. Therefore, there was insufficient data to pool results for Apley's and Thessaly at 5° into a meta-analysis. All nine studies were included in the meta-analysis to some degree (see table 3).

Three special tests were included in the meta-analysis: McMurray's,¹⁶ JLT¹⁰ and Thessaly at 20° knee flexion¹¹ (table 3). McMurray's had a pooled sensitivity of 61% (95% CI 45% to 74%) and a pooled specificity of 84% (95% CI 69% to 92%). JLT had a pooled sensitivity of 83% (95% CI 73% to 90%) and a pooled specificity of 83% (95% CI 61% to 94%). Thessaly 20° had a pooled sensitivity of 75% (95% CI 53% to 89%) and a pooled specificity of 87% (95% CI 65% to 96%).

LR+ of 3.2, 4.0 and 5.6, and LR– of 0.52, 0.23 and 0.28 for McMurray's, JLT and Thessaly 20° (see online supplementary figures S1–3), respectively. LR+ of between 0.2 and 0.5 indicate only small shifts in probability post-test.³⁴

Table 2 Included studies characteristic summary

Study	Mean age (range)	Mean symptom duration	Number and sex	Criterion standard	Affected meniscus	Test(s)	SN	SP	LR+	LR–	Authors' conclusion
Akseki <i>et al</i> ²⁹	35.7 (17–73)	32.4 months	110m 40F	Arthroscopy	Med Lat Med Lat	McMurray's JLT	0.67 0.53 0.88 0.67	0.69 0.88 0.44 0.80	2.16 4.42 1.57 3.35	0.48 0.53 0.27 0.41	Higher positive predictive values were obtained with McMurray's, but similar negative predictive values were achieved
Corea <i>et al</i> ⁶⁷	25.3 (18–40)	Not stated	93 sex not stated	Arthroscopy/ arthrotomy	Med Lat	McMurray's	0.65 0.52	0.93 0.94	9.51 7.94	0.38 0.52	McMurray's test is of poor sensitivity for detecting injured menisci. A negative test, however, is useful in excluding a meniscal tear
Eren ³⁰	19.2 (18–20)	14 months	104M	Arthroscopy	Med Lat	JLT	0.86 0.92	0.67 0.97	2.61 30.67	0.21 0.08	JLT as a test for lateral meniscal tears is accurate; however, for medial meniscal tears, rates are lower
Galli <i>et al</i> ³¹	29.7 (SD ±11.6)	52 months	39M 17F	Arthroscopy	Med/Lat Med/Lat	McMurray's JLT	0.34 0.63	0.86 0.50	2.52 1.26	0.76 0.74	JLT alone is of little clinical usefulness. A negative McMurray test does not modify the pretest probability of a meniscal lesion, while a positive result has a fair predictive value
Karachalios <i>et al</i> ¹¹	29.9 (18–56)	Not stated	301M 109F	MRI	Med Lat Med Lat Med Lat Med Lat Med Lat	McMurray's Apley's JLT Thessaly 5° Thessaly 20°	0.48 0.65 0.41 0.41 0.71 0.78 0.66 0.81 0.89 0.92	0.94 0.86 0.93 0.86 0.87 0.90 0.96 0.91 0.97 0.96	8.00 4.64 5.86 2.93 5.46 7.80 16.50 9.00 29.67 23.00	0.55 0.41 0.63 0.69 0.33 0.24 0.35 0.21 0.11 0.08	The Thessaly test at 20° of knee flexion can be used effectively as a first-line clinical screening test for meniscal tears
Konan <i>et al</i> ⁵³	39 (16–56)	Not stated	80M 29F	Arthroscopy	Med Lat Med Lat Med Lat	McMurray's JLT Thessaly 5°	0.50 0.21 0.83 0.68 0.41 0.16	0.77 0.94 0.76 0.97 0.68 0.89	2.17 3.50 3.46 22.67 1.29 1.44	0.65 0.84 0.22 0.33 0.86 0.95	Physical tests may not always be diagnostic of meniscal tears. MRI and arthroscopy may be essential in dubious clinical presentations and especially where more than one pathology is suspected
Manzotti <i>et al</i> ⁶⁸	32.4 (17–48)	26.5 months	94M 36F	Arthroscopy	Med Lat	McMurray's	0.88 0.79	0.50 0.20	1.76 0.99	0.24 1.50	McMurray's test is a valuable diagnostic tool when used in association with other clinical manoeuvres
Mirzatolooie <i>et al</i> ⁶⁹	26.63 (17–40)	Not stated	76M 4F	Arthroscopy	Med/Lat Med/Lat Med/Lat	McMurray's JLT Thessaly 20°	0.51 0.92 0.79	0.91 0.63 0.40	5.67 2.49 1.32	0.54 0.13 0.53	The Thessaly test has a low specificity in patients with combined anterior cruciate ligament and meniscal injuries and cannot be recommended as a diagnostic test in this setting
Rinonapoli <i>et al</i> ²¹	27.8 (20–50)	Not stated	75M 27F	Arthroscopy	Med/Lat Med/Lat	McMurray's Apley's	0.80 0.84	0.79 0.71	3.71 2.93	0.26 0.23	The present study revealed that the assessed clinical tests for detecting meniscal tears are only partially reliable

F, female; JLT, joint line tenderness; Lat, lateral; LR–, negative likelihood ratio; LR+, positive likelihood ratio; M, male; Med, medial; SN, sensitivity; SP, specificity.

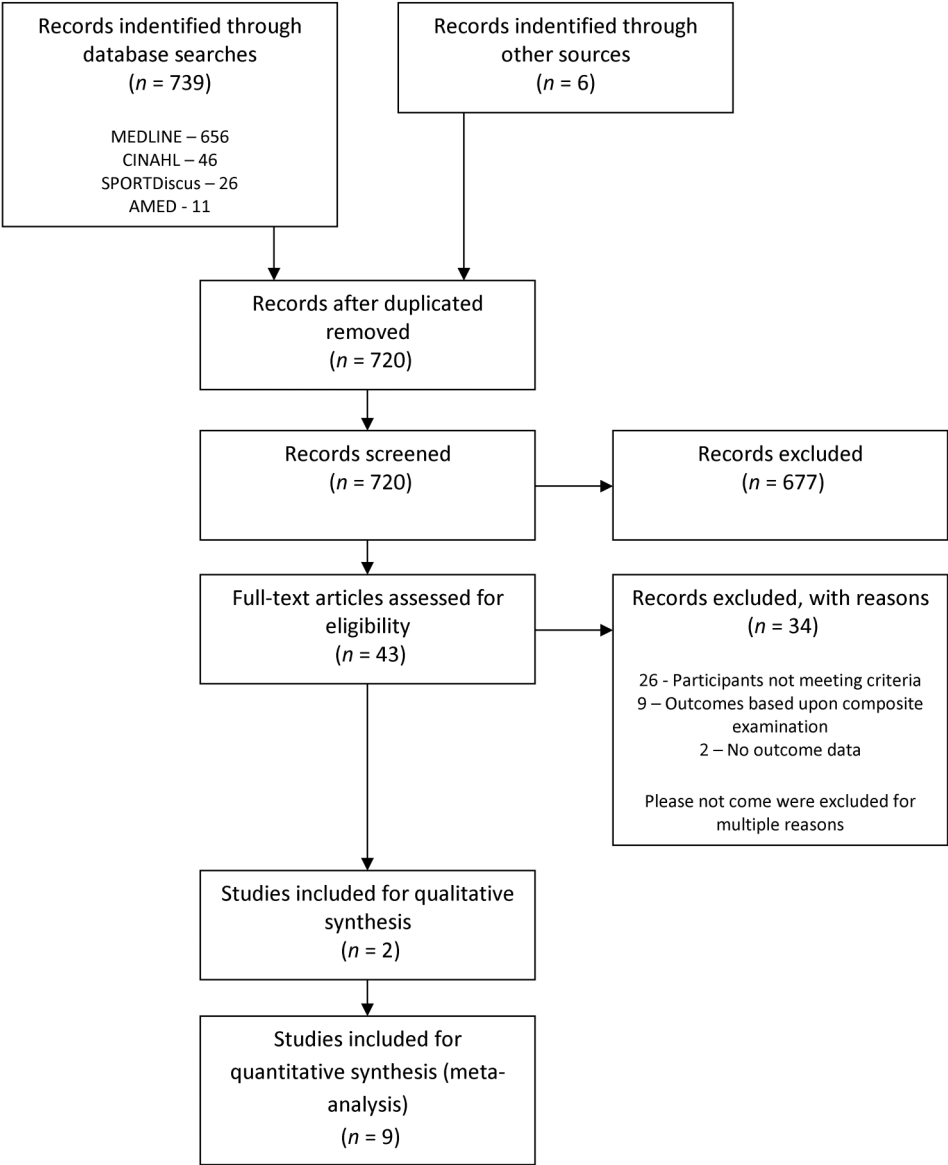


Figure 1 Study selection process (CINAHL, Cumulative Index to Nursing and Allied Health Literature; AMED, The Allied and Complementary Medicine Database).

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Akseki 2004							
Corea 1994							
Eren 2003							
Galli 2013							
Karachalios 2005							
Konan 2009							
Manzotti 1997							
Mirzatoioei 2010							
Rinonapoli 2011							
Low Risk High Risk Unclear Risk							

Figure 2 Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS) 2 score breakdown.

Table 3 Summary of sensitivity, specificity, likelihood ratios and heterogeneity

	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	I ² (%)
McMurray's	61% (45% to 74%)	84% (69% to 92%)	3.2 (1.7 to 5.9)	0.52 (0.34 to 0.81)	51
JLT	83% (73% to 90%)	83% (61% to 94%)	4.0 (2.1 to 7.5)	0.23 (0.12 to 0.44)	83
Thessaly 20°	75% (53% to 89%)	87% (65% to 96%)	5.6 (1.5 to 21.0)	0.28 (0.11 to 0.71)	94

JLT, joint line tenderness; LR-, negative likelihood ratio; LR+, positive likelihood ratio.

Two of the tests, JLT and Thessaly 20°, had a high heterogeneity I² score, with McMurray's having a moderate between-study heterogeneity I² score. These data, coupled with the relatively low shifts in probability with the likelihood ratios,⁷⁰ show that the three tests analysed will not accurately diagnose a torn meniscus.

Apley's test had a combined (medial and lateral) sensitivity of 84% and 20% and specificity of 79% and 84% (Rinonapoli *et al*²¹ and Karachalios *et al*,¹¹ respectively). Thessaly 5° had a combined (medial and lateral) sensitivity of 35% and 65% and specificity of 89% and 82% (Konan *et al*⁵³ and Karachalios *et al*,¹¹ respectively).

Discussion

The main objective of this systematic review was to synthesise the most up-to-date literature for diagnostic accuracy studies for meniscal tears of the knee for adults. The overall results of the three tests that were included within the meta-analysis (McMurray's, JLT and Thessaly 20°) indicate that they have poor accuracy. The pooled meta-analyses indicate that McMurray's will diagnose 61% of people presenting with a meniscal tear, Thessaly better at 75% and JLT best at 83%. False-positive findings are likely to be approximately 20% for all three tests. However, these results should be used with caution, due to the low number of included studies, poor quality of the studies and high levels of heterogeneity.

Apley's test, not included within the meta-analysis, was investigated by two studies. Performance varied considerably between the two; for example, sensitivity varied from 84%²¹ to 41%.¹¹ One possible cause for this difference is that Karachalios *et al*¹¹ included participants with no knee symptoms, and Rinonapoli *et al*²¹ suffered from verification bias, which could overestimate the sensitivity of a test, since prevalence within the sample size is larger.⁷¹

Combined lateral and medial sensitivity of the Thessaly 5° test varied from 35%⁵³ to 65%.¹¹ Reasons for this are unclear as patient selection and exclusion criteria were similar in both studies. One explanation could be the different reference standards used. Karachalios *et al*'s¹¹ study was the only study that used MRI. Karachalios *et al*¹¹ also developed the Thessaly test, and therefore may have interpreted results differently or may have biased results inadvertently towards their own test. Karachalios *et al*¹¹ gives the Thessaly test substantially higher sensitivity and specificity scores than all the other studies that investigated it and their study was one of those that failed to supply for raw 2x2 data for meta-analysis and data checking despite being contacted for these data.

One study subcategorised results by concomitant injuries (anterior cruciate ligament tears) and found that

the accuracy of the individual tests (JTL, McMurray's and Thessaly) was lower.⁵³

Limitations of included studies

For the meta-analysis of McMurray's, JLT and Thessaly 20°, moderate to high levels of heterogeneity existed. I² scores were 51%, 83% and 94%, respectively. This reduced the robustness and usefulness of the pooled data. In general, wide variation in test procedures were applied to a wide variety of patients including different ages, sex ratios and duration of symptoms. Despite trying to limit heterogeneity by excluding studies with children, there still existed a wide range of ages and sexes. One study, for example, only had participants between the ages of 18 and 20,³⁰ while the majority of studies had participants up to 40 and 50 years of age, with one study's eldest participant being 73 years old.²⁹

There was also wide variation in how the special tests were performed (see table 1). Considering that there are poor levels of inter-rater reliability found with McMurray's test when examiners have agreed on the test procedure,^{40 72} it is plausible that this accounts for the majority of the heterogeneity.

Another possible cause of heterogeneity between included studies is the differing prevalence rates within each sample. Verification bias can exaggerate the prevalence of the disease within the sample and, as a consequence, overestimate the sensitivity and underestimate the specificity.⁷¹ The prevalence rates of meniscal tears (as confirmed by the reference test) across the three tests included within the meta-analysis varied hugely. For example, prevalence for McMurray's ranged from 35%⁶⁷ to 88%.⁶⁸ Prevalence for JLT ranged from 31%³⁰ to 64%³¹ and Thessaly 20° ranged from 21%¹¹ to 49%.^{53 69} Another limitation of the included studies is that all but one study used arthroscopy as the gold standard test, and it is thought that this also introduces verification bias.

Although all studies scored 'low risk' in total on the QUADAS 2 tool, no study received a score of low risk in all categories for the risk of bias. The main methodological errors were with index test description, no confirmation of blinding for the reference test, poor description of flow and timing, poor details given with regard to dropouts and the number of patients being included within the analysis.

Limitations of this review

An extensive literature search was carried out. To reduce risk of bias, two reviewers screened full texts independently for inclusion. An attempt was made to source unpublished trials; however, it is possible that not all publications were retrieved. Furthermore, language bias

remained, since no attempt was made to source studies published in any other language than English.

Other limitations of this review are that dichotomised subgroup analysis by prevalence rates and clinical test definitions were not carried out. This may have reduced the between-study heterogeneity and improved robustness of the data synthesis. However, as many of the included studies lacked a clear definition and/or contradicted themselves, this was not considered possible.

Statistical pooling of data for sensitivity and specificity may not represent an accurate estimate, and clinicians should be aware of this and interpret with caution.

Our inter-rater agreement of the QUADAS 2 scores was only moderate to fair.^{25 26} However, as the main analysis was carried out with all included studies and the QUADAS 2 tool was not used to subcategorise studies, these difficulties would not have affected the conclusions of this review.

Comparison with other reviews

Despite updating the data with more narrow inclusion criteria, and including four more studies,^{21 31 53 69} our main findings differ very little from the previous four systematic reviews that included a meta-analysis^{9 12 13 15} (see online supplementary table S2).

The robustness of our results and conclusion are greater than those of the other systematic reviews. Our study was the first to use the QUADAS 2 tool, the first to perform a meta-analysis for Thessaly's test and the first to limit our search to adults only. Furthermore, the methodological quality of this review is guided by the PRISMA statement.²²

Clinical and research implications

It is known that levels of pain do not correlate to the presence of meniscal tears of the knee,^{7 8 73} and that both peripheral and central sensitisation can be an underlying mechanism for people with chronic knee pain,^{74–76} and musculoskeletal pain in general.⁷⁴ A recent systematic review showed that arthroscopy meniscectomy for degenerative meniscal tears works no better than sham/placebo surgery or versus conservative treatment.⁷⁷ This starts to question the clinical need for such a diagnosis, since it is perceivable that 'confirmed' meniscal tears on MRI may be incidental, and therefore play no part in the development of pain or loss of function. Given that the prevalence of meniscal tears increases with age, and is almost double in people with radiological evidence of osteoarthritic changes,⁷³ the accuracy may be different in different age groups, but the usefulness of this in relation to their management is questionable. The incidental MRI findings in the spine are almost as common as meniscal tears of the knee.^{73 78} It is thought that these incidental spinal findings may actually have adverse effects for the patient leading to long-term fear avoidance.⁷⁹ No study has been found that looks at the implications of this for patients presenting with meniscal tears, but it is perceivable that false-positive findings could increase fear avoidance and limit restoration of normal knee function.

Primary clinicians must still be aware of the need to recognise if conservative treatment is not appropriate,

such as in cases of a 'locked knee' or true giving way.⁸⁰ A 'mechanical' block to a full range of movement or a 'mechanically' unstable knee would usually indicate an MRI and a surgical opinion.⁸¹ In these situations, the principal aim of the assessment is not to diagnose the specific tissue structure at fault, but to identify patients who are not appropriate for conservative rehabilitation in a timely manner, and the special tests would not support the clinical decision-making for this. It is likely that this would hold true for acute and degenerative tears.

In clinical practice, the tests are often not used in isolation alone, but are frequently used in combination with each other. Using the tests this way may produce a more accurate diagnosis; however, this cannot be confirmed or concluded from the current data. It is thought by the current authors that clinicians should abandon the tests that are based on the pathological model that lacks validity and reliability. Future research should focus on identifying which clinical characteristics might be useful either as prognostic indicators or have management implications for identifying conservative or surgical management.

Conclusion

The results of this systematic review indicate that the accuracy of McMurray's, Apley's, JLT and Thessaly to diagnose meniscal tears remains poor. This conclusion must be taken with caution since frequent methodological design flaws exist within the included studies, most studies suffered from various biases, and between-study heterogeneity makes pooled data unreliable.

The latest research surrounding meniscal tears within asymptomatic patients, and modern thinking with regard to pain and lack of efficacy for surgical treatment starts to challenge the need for such a diagnosis and use of special tests. Having a diagnosis of a meniscal tear is unlikely to help with the rehabilitation process and may induce fear avoidance.

This review cannot recommend the use of special tests for diagnosing meniscal tears. It is unclear if further research would considerably alter this conclusion.

Twitter Follow Benjamin Smith at @benedsmith

Acknowledgements The authors are grateful to Arthritis Research UK and the Chartered Society of Physiotherapy Charitable Trust for providing the funding for BES to complete this study.

Contributors BES was responsible for conception and design, publication screening, acquisition of data, analysis and interpretation, as well as drafting and revising the manuscript. DT was responsible for publication screening, analysis, reviewing and revising the manuscript. AC was responsible for acquisition of data, as well as reviewing and revising the manuscript. MH was responsible for reviewing and revising the manuscript. All authors have read and approved the final manuscript.

Competing interests None.

Provenance and peer review Not commissioned;
externally peer reviewed.

References

- Baker P, Reading I, Cooper C, *et al.* Knee disorders in the general population and their relation to occupation. *Occup Environ Med* 2003;60:794–7.
- O'Reilly SC, Muir KR, Doherty M. Occupation and knee pain: a community study. *Osteoarthritis Cartilage* 2000;8:78–81.
- Nuffield Health. A quarter of Britons suffering knee joint pain. 2012. <http://www.nuffieldhealth.com/hospitals/news/nuffield-health-knee-care-20-11-12>
- Snoeker BAM, Bakker EWP, Kegel CAT, *et al.* Risk factors for meniscal tears: a systematic review including meta-analysis. *J Orthop Sports Phys Ther* 2013;43:352–67.
- Draijer LW, Belo JN, Berg HF, *et al.* [Summary of the practice guideline "Traumatic knee problems" (first revision) from the Dutch College of General Practitioners]. *Ned Tijdschr Geneesk* 2010;154:A2225. <http://www.ncbi.nlm.nih.gov/pubmed/21429259> (accessed 6 Aug 2014).
- Baker P, Coggon D, Reading I, *et al.* Sports injury, occupational physical activity, joint laxity, and meniscal damage. *J Rheumatol* 2002;29:557–63. <http://www.ncbi.nlm.nih.gov/pubmed/11908573> (accessed 6 Aug 2014).
- Zanetti M, Pfirrmann CWA, Schmid MR, *et al.* Patients with suspected meniscal tears: prevalence of abnormalities seen on MRI of 100 symptomatic and 100 contralateral asymptomatic knees. *AJR Am J Roentgenol* 2003;181:635–41.
- Guermazi A, Niu J, Hayashi D, *et al.* Prevalence of abnormalities in knees detected by MRI in adults without knee osteoarthritis: population based observational study (Framingham Osteoarthritis Study). *BMJ* 2012;345:e5339.
- Hegedus E, Cook C, Hasselblad V, *et al.* Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. *J Orthop Sports Phys Ther* 2007;37:541–50.
- Malanga GA, Andrus S, Nadler SF, *et al.* 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.
- Karachalios T, Hantes M, Zibis AH, *et al.* Diagnostic accuracy of a new clinical test (the Thessaly test) for early detection of meniscal tears. *J Bone Joint Surg Am* 2005;87:955–62.
- Scholten RJ, Devillé WL, Opstelten W, *et al.* The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis. *J Fam Pract* 2001;50:938–44.
- Solomon DH, Simel DL, Bates DW, *et al.* The rational clinical examination. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. *JAMA* 2001;286:1610–20.
- Ryzewicz M, Peterson B, Siparsky PN, *et al.* The diagnosis of meniscus tears: the role of MRI and clinical examination. *Clin Orthop Relat Res* 2007;455:123–33.
- Meserve BB, Cleland JA, Boucher TR. A meta-analysis examining clinical test utilities for assessing meniscal injury. *Clin Rehabil* 2008;22:143–61.
- McMurray TP. The semilunar cartilages. *Br J Surg* 1942;29:407–14.
- Hing W, White S, Reid D, *et al.* Validity of the McMurray's test and modified versions of the test: a systematic literature review. *J Man Manip Ther* 2009;17:22–35. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2704345&tool=pmcentrez&rendertype=abstract> (accessed 5 Aug 2014).
- Apley G. The diagnosis of meniscus injuries some new clinical methods. *J Bone Joint Surg* 1947;29:78–84. <http://jbj.s.org/article.aspx?articleid=10415> (accessed 29 Dec 2013).
- Tria AJ. Clinical examination of the knee. In: Insall JN, Scott WN, eds. *Surgery of the knee*. Vol 1. 3rd edn. New York: Churchill Livingstone, 2001:161–74.
- Goldby LJ, Moore AP, Doust J, *et al.* A randomized controlled trial investigating the efficiency of musculoskeletal physiotherapy on chronic low back disorder. *Spine (Phila Pa 1976)* 2006;31:1083–93.
- Rinonapoli G, Carraro A, Delcogliano A. The clinical diagnosis of meniscal tear is not easy. Reliability of two clinical meniscal tests and magnetic resonance imaging. *Int J Immunopathol Pharmacol* 2011;24:39–44.
- Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
- Harbord RM, Deeks JJ, Egger M, *et al.* A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239–51.
- Reitsma JB, Glas AS, Rutjes AWS, *et al.* Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Fleiss JL. *Statistical methods for rates and proportions*. 2nd edn. New York: John Wiley, 1981.
- Higgins J, Deeks J. Cochrane handbook: general methods for Cochrane reviews: Ch 7: selecting studies and collecting data. In: Higgins PTJ, Green S, eds. *Cochrane handbook for: systematic reviews of interventions*. Wiley-Blackwell, 2011:151–86.
- Wallace BC, Dahabreh IJ, Trikalinos TA, *et al.* Closing the gap between methodologists and end-users: R as a computational back-end. *Wiley Interdiscip Rev Comput* 2012;49:1–15.
- Akseki D, Ozcan O, Boya H, *et al.* A new weight-bearing meniscal test and a comparison with McMurray's test and joint line tenderness. *Arthroscopy* 2004;20:951–8.
- Eren OT. The accuracy of joint line tenderness by physical examination in the diagnosis of meniscal tears. *Arthroscopy* 2003;19:850–4.
- Galli M, Ciriello V, Menghi A, *et al.* Joint line tenderness and McMurray tests for the detection of meniscal lesions: what is their real diagnostic value? *Arch Phys Med Rehabil* 2013;94:1126–31.
- Whiting P, Rutjes A, Westwood M, *et al.* QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
- Higgins J, Thompson S, Deeks J, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
- Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703–7. <http://www.ncbi.nlm.nih.gov/pubmed/8309035> (accessed 5 Aug 2014).
- Sackett D, Tugwell P, Guyatt G. *Clinical epidemiology: a basic science for clinical medicine*. 2nd edn. Lippincott Williams and Wilkins, 1991.
- Anderson AF, Lipscomb AB. Clinical diagnosis of meniscal tears. Description of a new manipulative test. *Am J Sports Med* 1986;14:291–3. <http://www.ncbi.nlm.nih.gov/pubmed/3755297> (accessed 17 Jul 2014).
- Barry OCD, Smith H, McManus F, *et al.* Clinical assessment of suspected meniscal tears. *Ir J Med Sci* 1983;152:149–51.
- Boeree NR, Ackroyd CE. Assessment of the menisci and cruciate ligaments: an audit of clinical practice. *Injury* 1991;22:291–4.
- Curtin W, O'Farrell D, McGoldrick F, *et al.* The correlation between clinical diagnosis of knee pathology and findings at arthroscopy. *Ir J Med Sci* 1992;161:135–6.
- Evans PJ, Bell GD, Frank C. Prospective evaluation of the McMurray test. *Am J Sports Med* 1993;21:604–8. <http://www.ncbi.nlm.nih.gov/pubmed/8368424> (accessed 17 Jul 2014).

41. Fowler PJ, Lubliner JA. The predictive value of five clinical signs in the evaluation of meniscal pathology. *Arthroscopy* 1989;5:184–6. <http://www.ncbi.nlm.nih.gov/pubmed/2775390>
42. Harrison BK, Abell BE, Gibson TW. The Thessaly test for detection of meniscal tears: validation of a new physical examination technique for primary care medicine. *Clin J Sport Med* 2009;19:9–12.
43. Johnson LL, Johnson AL, Colquitt JA, *et al.* Is it possible to make an accurate diagnosis based only on a medical history? A pilot study on women's knee joints. *Arthroscopy* 1996;12:709–14. <http://www.ncbi.nlm.nih.gov/pubmed/9115560>
44. Kim S-J, Hwang B-Y, Choi D-H, *et al.* The paradoxical McMurray test for the detection of meniscal tears: an arthroscopic study of mechanisms, types, and accuracy. *J Bone Joint Surg Am* 2012;94:e1181–7.
45. Kocabey Y, Tetik O, Isbell WM, *et al.* The value of clinical examination versus magnetic resonance imaging in the diagnosis of meniscal tears and anterior cruciate ligament rupture. *Arthroscopy* 2004;20:696–700.
46. Kurosaka M, Yagi M, Yoshiya S, *et al.* Efficacy of the axially loaded pivot shift test for the diagnosis of a meniscal tear. *Int Orthop* 1999;23:271–4. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3619761&tool=pmcentrez&rendertype=abstract>
47. Mariani PP, Adriani E, Maresca G, *et al.* A prospective evaluation of a test for lateral meniscus tears. *Knee Surg Sports Traumatol Arthrosc* 1996;4:22–6. <http://www.ncbi.nlm.nih.gov/pubmed/8819059>
48. Miller GK. A prospective study comparing the accuracy of the clinical diagnosis of meniscus tear with magnetic resonance imaging and its effect on clinical outcome. *Arthroscopy* 1996;12:406–13. <http://www.ncbi.nlm.nih.gov/pubmed/8863997>
49. Muellner T, Weinstabl R, Schabus R, *et al.* The diagnosis of meniscal tears in athletes. A comparison of clinical and magnetic resonance imaging investigations. *Am J Sports Med* 1997;25:7–12. <http://www.ncbi.nlm.nih.gov/pubmed/9006685> (accessed 17 Jul 2014).
50. Nickinson R, Darrah C, Donell S. Accuracy of clinical diagnosis in patients undergoing knee arthroscopy. *Int Orthop* 2010;34:39–44.
51. Noble J, Erat K. In defence of the meniscus. A prospective study of 200 meniscectomy patients. *J Bone Joint Surg Br* 1980;62-B:7–11. <http://www.ncbi.nlm.nih.gov/pubmed/7351438> (accessed 17 Jul 2014).
52. Pookarnjanamorakot C, Korsantirat T, Woratanarat P. Meniscal lesions in the anterior cruciate insufficient knee: the accuracy of clinical evaluation. *J Med Assoc Thai* 2004;87:618–23. <http://www.ncbi.nlm.nih.gov/pubmed/15279338>
53. Konan S, Rayan F, Haddad FS. Do physical diagnostic tests accurately detect meniscal tears? *Knee Surg Sports Traumatol Arthrosc* 2009;17:806–11.
54. Rose REC. The accuracy of joint line tenderness in the diagnosis of meniscal tears. *West Indian Med J* 2006;55:323–6. <http://www.ncbi.nlm.nih.gov/pubmed/17373299> (accessed 17 Jul 2014).
55. Saengnipanthkul S, Sirichativapee W, Kowsuwon W, *et al.* The effects of medial patellar plica on clinical diagnosis of medial meniscal lesion. *J Med Assoc Thai* 1992;75:704–8. <http://www.ncbi.nlm.nih.gov/pubmed/1308539> (accessed 17 Jul 2014).
56. Shelbourne KD, Benner RW. Correlation of joint line tenderness and meniscus pathology in patients with subacute and chronic anterior cruciate ligament injuries. *J Knee Surg* 2009;22:187–90. <http://www.ncbi.nlm.nih.gov/pubmed/19634720>
57. Simonsen O, Jensen J, Mouritsen P, *et al.* The accuracy of clinical examination of injury of the knee joint. *Injury* 1984;16:96–101. <http://www.ncbi.nlm.nih.gov/pubmed/6547924>
58. Yan R, Wang H, Yang Z, *et al.* Predicted probability of meniscus tears: comparing history and physical examination with MRI. *Swiss Med Wkly* 2011;141:w13314.
59. Rose N, 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.
60. Shelbourne KD, Martini DJ, McCarroll JR, *et al.* Correlation of joint line tenderness and meniscal lesions in patients with acute anterior cruciate ligament tears. *Am J Sports Med* 1995;23:166–9. <http://www.ncbi.nlm.nih.gov/pubmed/7778700>
61. Sladjan T, Zoran V, Zoran B. Correlation of clinical examination, ultrasound sonography, and magnetic resonance imaging findings with arthroscopic findings in relation to acute and chronic lateral meniscus injuries. *J Orthop Sci* 2014;19:71–6. <http://offcampus.lib.washington.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=93922694&site=ehost-live>
62. Esmaili Jah AA, Keyhani S, Zarei R, *et al.* Accuracy of MRI in comparison with clinical and arthroscopic findings in ligamentous and meniscal injuries of the knee. *Acta Orthop Belg* 2005;71:189–96. <http://www.ncbi.nlm.nih.gov/pubmed/16152853>
63. Sharma UK, Shrestha BK, Rijal S, *et al.* Clinical, MRI and arthroscopic correlation in internal derangement of knee. *Kathmandu Univ Med J (KUMJ)* 2011;9:174–8. <http://www.ncbi.nlm.nih.gov/pubmed/22609502>
64. Wagemakers HP, Heintjes EM, Boks SS, *et al.* Diagnostic value of history-taking and physical examination for assessing meniscal tears of the knee in general practice. *Clin J Sport Med* 2008;18:24–30.
65. Abdon P, Lindstrand A, Thorngren K-G. Statistical evaluation of the diagnostic criteria for meniscal tears. *Int Orthop* 1990;14:341–5.
66. Ercin E, Kaya I, Sungur I, *et al.* History, clinical findings, magnetic resonance imaging, and arthroscopic correlation in meniscal lesions. *Knee Surg Sports Traumatol Arthrosc* 2012;20:851–6.
67. Corea JR, Moussa M, al Othman A. McMurray's test tested. *Knee Surg Sports Traumatol Arthrosc* 1994;2:70–2. <http://www.ncbi.nlm.nih.gov/pubmed/7584186> (accessed 17 Jul 2014).
68. Manzotti A, Baiguini P, Locatelli A, *et al.* Statistical evaluation of McMurray's test in the clinical diagnosis of meniscus injuries. *J Sport Traumatol Relat Res* 1997;19:83–9.
69. Mirzatooleei F, Yekta Z, Bayazidchi M, *et al.* Validation of the Thessaly test for detecting meniscal tears in anterior cruciate deficient knees. *Knee* 2010;17:221–3.
70. McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002;17:647–50.
71. Zhou XH. Correcting for verification bias in studies of a diagnostic test's accuracy. *Stat Methods Med Res* 1998;7:337–53.
72. Dervin GF, Stiell IG, Wells GA, *et al.* Physicians' accuracy and interrater reliability for the diagnosis of unstable meniscal tears in patients having osteoarthritis of the knee. *Can J Surg* 2001;44:267–74. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3692659&tool=pmcentrez&rendertype=abstract> (accessed 22 Jul 2014).
73. Englund M, Guermazi A, Gale D, *et al.* Incidental meniscal findings on knee MRI in middle-aged and elderly persons. *N Engl J Med* 2008;359:1108–15.
74. Bajaj P, Graven-Nielsen T, Arendt-Nielsen L. Osteoarthritis and its association with muscle hyperalgesia: an experimental controlled study. *Pain* 2001;93:107–14. <http://www.ncbi.nlm.nih.gov/pubmed/11427321> (accessed 24 Jul 2014).
75. Imamura M, Imamura ST, Kaziya HHS, *et al.* Impact of nervous system hyperalgesia on pain, disability, and quality of life in patients with knee osteoarthritis: a controlled analysis. *Arthritis Rheum* 2008;59:1424–31.
76. Kosek E, Ordeberg G. Lack of pressure pain modulation by heterotopic noxious conditioning stimulation in patients with painful osteoarthritis before, but not following, surgical pain

- relief. *Pain* 2000;88:69–78. <http://www.ncbi.nlm.nih.gov/pubmed/11098101> (accessed 24 Jul 2014).
77. Khan M, Evaniew N, Bedi A, *et al.* Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. *CMAJ* 2014;186:1057–64.
 78. Boden SD, McCowin PR, Davis DO, *et al.* Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg* 1990;72:1178–84. <http://www.ncbi.nlm.nih.gov/pubmed/2398088>
 79. Chou R, Qaseem A, Snow V, *et al.* Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478–91.
 80. Nicholas SJ, Golant A, Schachter AK, *et al.* A new surgical technique for arthroscopic repair of the meniscus root tear. *Knee Surg Sports Traumatol Arthrosc* 2009;17:1433–6.
 81. Brukner P, Khan K. *Clinical sports medicine*. McGraw-Hill Medical, 2012.