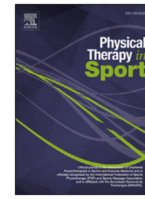
Physical Therapy in Sport 23 (2017) 143–155

Contents lists available at ScienceDirect

Physical Therapy in Sport

journal homepage: www.elsevier.com/ptsp



Diagnostic validity of physical examination tests for common knee disorders: An overview of systematic reviews and meta-analysis

Simon Décary [a, b], Philippe Ouellet [a, b], Pascal-André Vendittoli [b, c], Jean-Sébastien Roy [d, e], François Desmeules [a, b, *]

[a] School of Rehabilitation, Faculty of Medicine, University of Montreal, Montreal, Quebec, Canada
[b] Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, University of Montreal Affiliated Research Center, Montreal, Quebec, Canada
[c] Department of Surgery, Maisonneuve-Rosemont Hospital, University of Montreal, Montreal, Quebec, Canada
[d] Department of Rehabilitation, Faculty of Medicine, Laval University, Quebec City, Quebec, Canada
[e] Centers for Interdisciplinary Research in Rehabilitation and Social Integration, Quebec City, Canada

A R T I C L E I N F O

Article history:
Received 30 March 2016
Received in revised form
2 August 2016
Accepted 4 August 2016

Keywords:
Knee
Validity
Physical examination tests
Diagnostic

A B S T R A C T

Introduction: More evidence on diagnostic validity of physical examination tests for knee disorders is needed to lower frequently used and costly imaging tests.
Objective: To conduct a systematic review of systematic reviews (SR) and meta-analyses (MA) evaluating the diagnostic validity of physical examination tests for knee disorders.
Methods: A structured literature search was conducted in five databases until January 2016. Methodological quality was assessed using the AMSTAR.
Results: Seventeen reviews were included with mean AMSTAR score of 5.5 ± 2.3. Based on six SR, only the Lachman test for ACL injuries is diagnostically valid when individually performed (Likelihood ratio (LR+):10.2, LR-:0.2). Based on two SR, the *Ottawa Knee Rule* is a valid screening tool for knee fractures (LR-:0.05). Based on one SR, the EULAR criteria had a post-test probability of 99% for the diagnosis of knee osteoarthritis. Based on two SR, a complete physical examination performed by a trained health provider was found to be diagnostically valid for ACL, PCL and meniscal injuries as well as for cartilage lesions.
Conclusion: When individually performed, common physical tests are rarely able to rule in or rule out a specific knee disorder, except the Lachman for ACL injuries. There is low-quality evidence concerning the validity of combining history elements and physical tests.

1. Introduction

Knee disorders and injuries are common reasons for consultation in primary care (Mody & Brooks, 2012). The lifetime prevalence of knee pain is 45%, and at least 31% of the affected individuals will consult a health care practitioner (Baker et al., 2003; Jette & Delitto, 1997). Common knee disorders include traumatic injuries such as meniscal injuries (Snoeker et al., 2013), anterior cruciate ligament (ACL) injuries (Granan et al., 2013), fractures (Bharam, Vrahas, & Fu, 2002) or overuse or degenerative disorders like osteoarthritis (Tveit et al., 2012), patellofemoral pain (PFP) (Barber Foss et al., 2012) and tendinopathies (Andersen et al., 2013). Knee disorders often result in disabilities as well as in a decrease in health-related quality of life and may lead to workplace absenteeism (Filbay et al., 2015; Laslett et al., 2012). Efficient management of patients suffering from knee disorders is often lacking because the initial diagnosis is either erroneous or incomplete. Too often clinicians rely on medical imaging which in turn increase healthcare costs and may incur unnecessary delays in diagnosis and initiation of care (Stein, 2005).

Abbreviations: SR, Systematic-review; MA, Meta-analysis; ACL, Anterior cruciate ligament injuries; PCL, Posterior cruciate ligament injuries; OA, Osteoarthritis; PFP, Patellofemoral pain; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Sn, Sensitivity; Sp, Specificity; PPV, Positive predictive value; NPV, Negative predictive value; LR, Likelihood ratio; DOR, Diagnostic odds ratio; AMSTAR, Assessment of methodological quality of multiple systematic reviews.

* Corresponding author. Orthopaedic Clinical Research Unit, Maisonneuve-Rosemont Hospital Research Center, University of Montreal Affiliated Research Center, Montreal, Quebec, Canada.
E-mail addresses: simon.decary.1@umontreal.ca (S. Décary), philippe.ouellet.3@umontreal.ca (P. Ouellet), pa_vendittoli@hotmail.com (P.-A. Vendittoli), jean-sebastien.roy@fmed.ulaval.ca (J.-S. Roy), f.desmeules@umontreal.ca (F. Desmeules).

http://dx.doi.org/10.1016/j.ptsp.2016.08.002
1466-853X/© 2016 Elsevier Ltd. All rights reserved.

Moreover, evidence suggests that medical imaging may be less valid than a complete physical examination in a large proportion of cases (Jackson, O'Malley, & Kroenke, 2003).

Clinicians rely on thorough patient history elements and physical examination tests to make a diagnosis where the patient's responses and findings are combined to make a valid diagnosis. This process remains the cornerstone for optimal, fast and efficient management of patients with musculoskeletal disorders (Jackson et al., 2003). However, evidence indicates that the ability to make a valid diagnosis for common knee disorders based on a complete physical examination in primary care remains a challenge (Skou, Thomsen, & Simonsen, 2014). Therefore, it is important to synthesize the evidence of the diagnostic validity of physical examination to better help clinicians in making a valid diagnosis.

In recent years, systematic reviews of primary diagnostic studies for all diagnoses and tests for hip (Rahman et al., 2013; Reiman et al., 2013) and shoulder disorders (Hegedus et al., 2012) have been published. Also, many new systematic reviews (SR) and meta-analysis (MA) regarding the diagnosis of meniscal injuries (Hegedus et al., 2007; Hing et al., 2009; Meserve, Cleland, & Boucher, 2008; Ryzewicz et al., 2007), ACL injuries (Benjaminse, Gokeler, & van der Schans, 2006; van Eck et al., 2013), knee osteoarthritis (Zhang et al., 2010) and patellofemoral pain (PFP) (Cook et al., 2012) and have been published. However, evidence for knee disorders has not been synthesized in one clinically useful format. Therefore, the objective of this study is to provide updated information to clinicians working with individuals affected by knee disorders by systematically reviewing all the SR and MA reporting the diagnostic validity of physical examination tests performed individually or in combination with patient history elements for diagnosis of common knee disorders.

2. Methods

2.1. Study design

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used to guide the design of this systematic review (Liberati et al., 2009; Moher et al., 2009). The tool is a list of 27 items, which assesses the adequate transparent reporting of the results of a systematic review or a meta-analysis.

2.2. Literature search and study identification

A literature search was performed in five bibliographical databases: Pubmed, Medline, CINAHL, Embase, the Cochrane Database of Systematic reviews and Pedro using relevant and MESH based keywords. The keywords were adapted to the various databases, as demonstrated in Appendix 1. Databases were searched from their date of inception to January 2016. References lists of included studies and important textbooks on musculoskeletal diagnosis were also investigated to verify the completeness of the current search (Cook & Hegedus, 2011; Magee, 2014).

2.3. Data extraction and quality assessment

2.3.1. Study selection and data extraction

Each article, title and abstract were screened by two independent reviewers to determine eligibility. To be included, articles needed to 1- be a systematic review or a meta-analysis, 2- report on the diagnostic properties of at least one physical test for at least one knee disorder and 3- be written in English or French.

Data extraction of the selected SR/MA included: the study design (i.e: SR or MA), the examination physical tests evaluated to diagnose a specific knee disorder, the number of studies included in the SR, the number of studies pooled for a MA, and the total number of participants included in the review. The diagnostic properties of the clinical tests under study were extracted and included (when reported): sensitivity (Sn), specificity (Sp), positive and negative predictive values (PPV/NPV), positive and negative likelihood ratios (LR+/-) and diagnostic odds ratio (DOR (LR+/LR-)).

2.3.2. Assessment of methodological quality of the included systematic reviews (AMSTAR)

The methodological quality of the included SR and MA was assessed with the AMSTAR tool (Shea et al., 2007a). The AMSTAR is a validated and highly reliable tool aimed at assessing the quality of systematic reviews (Shea et al., 2007b, 2009. Each of the 11 methodological items are marked as “yes”, “no”, “cannot answer” (when unclear or insufficient information to answer) and “not applicable” (Shea et al., 2007a). Two raters independently assessed the methodological quality of each included review, then compared ratings and resolved any differences if present. In order to achieve consensus, a structured process was employed, where rechecking of the facts in the text was initially performed, followed by a discussion of the adherence to standards as well as the use of an independent third rater in case of persistent disagreement.

A total score out of 11 can be calculated adding the number of “yes” answers (Shea et al., 2007b, 2009). To objectively synthesize and formulate recommendations, the methodological quality was used to establish the strength of evidence. A review with an AMSTAR score $\geq$8/11 was considered of high-quality, between 5 and 7/11 was considered of moderate-quality and <5 was considered of low-quality. No systematic reviews or meta-analysis were excluded based on methodological quality.

2.4. Data analyses

The mean methodological score of each review was calculated. Cohen's kappa was used to calculate pre-consensus inter-rater agreement on individual methodological items of the AMSTAR tool. Overall results and related 95% confidence intervals for each diagnostic property statistic extracted from the reviews were directly extracted when pooled results were presented in the original papers. If no pooled results for a given diagnostic property statistic were presented, the range of estimates reported in the SR was extracted. For each relevant diagnostic test, the range of point estimates for Sn, Sp, LR $\pm$ and DOR (where applicable) across all included SR/MA was reported for a qualitative assessment of the evidence.

Sensitivity (Se) and specificity (Sp) relate respectively to the proportion of true positive and true negative when a test is performed and do not inform on the probability of a patient having a disorder if a test is positive or negative, while PPV/NPV inform on that probability considering the prevalence of the studied sample (Hing et al., 2009; Jaeschke, Guyatt, & Sackett, 1994). Because of these issues, the positive and negative likelihood ratio (calculated using both Se and Sp and representing the odds for a patient of having or not having a disorder) is advocated to guide clinical decision-making (Hegedus et al., 2007; Jaeschke et al., 1994). Although no universal agreement exists, a test was considered valid if it reaches a LR+ of 5 or more and a LR-of 0.2 or less in order to formulate recommendations regarding its validity (Jaeschke et al., 1994). These values produce a moderate shift in post-test probability, signifying that a disorder is present if the test is positive and absent if the test is negative, and when a test performs up to these values, it is useful to make a valid diagnosis (Cook et al., 2012; Jaeschke et al., 1994).

3. Results

3.1. Overall description of included reviews

As shown in Fig. 1, 6750 potential studies were initially identified, 6669 articles were excluded and 17 reviews were ultimately included. Table 1 presents the overall characteristics of the included reviews. Eleven reviews included a MA while the six others were SR without a MA. Eight reviews evaluated meniscal injuries, and the most common tests were the McMurray, the Apley's manoeuver and the joint line tenderness tests (Table 3). Six reviews assessed ACL injuries, and the most common tests evaluated were the Lachman, the anterior drawer and the pivot shift tests (Table 4). One clinical prediction rule for the screening of knee fractures was evaluated in two reviews (Table 5). One clinical prediction rule and one set of diagnostic criteria to diagnose knee osteoarthritis were the focus of two reviews (Table 6). Combinations of tests for ACL and posterior cruciate ligament (PCL) tears, meniscal injuries and for cartilage defects were appraised in two reviews (Table 7).

3.2. Methodological quality of included reviews

The AMSTAR scores for the assessment of methodological quality of the SR/MA are presented in Table 2. AMSTAR score ranged from 1 to 8 (out of 11) with a mean of 5.5 ± 2.3 indicating a moderate methodological quality. Seven reviews reached a score of 7 or higher. More than ten out of seventeen reviews performed a comprehensive literature search (item 3), provided a list of studies (item 5), provided characteristics of the included studies (item 6) and assessed the scientific quality of the included studies (item 7). Between seven and ten out of seventeen reviews duplicated study selection and data extraction (item 2), used the status of publication as an inclusion criteria (item 4), used the scientific quality of the included studies in the conclusion (item 8), used appropriate methods to combine findings (item 9) and stated potential conflict of interest (item 11). However, only one review assessed the likelihood of publication bias (item 10) while no reviews provided an “a priori” design (item 1). The average inter-rater reliability for individual item was substantial (kappa = 0.69). The majority of items reached substantial agreements ($\kappa > 0.6$) (items 1, 4, 5, 6, 7, 8, 9, 11). All other items reached a moderate agreement ($\kappa > 0.4$) (items 2, 3, 10). After discussion between the two raters, consensus was always achieved.

3.3. Summary of findings

3.3.1. Meniscal injuries

Table 3 presents the diagnostic properties of the physical tests used for the diagnosis of meniscal injuries. Eight SR/MA provided data on the diagnosis of meniscal injuries (Hegedus et al., 2007; Hing et al., 2009; Jackson et al., 2003; Meserve et al., 2008; Ryzewicz et al., 2007; Scholten et al., 2001; Smith et al., 2015; Solomon et al., 2001) with AMSTAR scores ranging from 2 to 8 out of 11 with a mean of 4.8. Based on the data extracted from the review by *Hegedus* et al. (Hegedus et al., 2007), the highest quality review with the most studies included in the meta-analyses (8/11, n = 18), the McMurray's test demonstrated the highest Sn with a score of 70.5% (95% CI: 67.4–73.4%) (Hegedus et al., 2007). The point estimates for Sn varied across all included SR/MA (range: 52.0–70.5%). Also based on the data by *Hegedus* et al., the joint line tenderness demonstrated the highest Sp with 77.4% (95% CI: 75.6–79.1%) (Hegedus et al., 2007). Again, the point estimates for the Sp of this test varied across all included SR/MA (range: 29.0–83.0%). Based on the data extracted from *Smith* et al. (Smith et al., 2015), the most recent and highest quality review to provide data for the LR+/- (8/11, n = 9), the joint line tenderness also demonstrated both the highest LR+ with 4.0 (95% CI: 2.1–7.5) and the lowest LR-with 0.23 (95% CI: 0.12–0.44). The point estimates of the joint line tenderness LR ± varied across others included SR/MA for both LR+/- (range LR+: 0.9–4.0; range LR-: 0.23–1.1). The Thessaly's test diagnostic validity was also presented for the first time by the review of *Smith* et al. (Smith et al., 2015), with a reported LR+ of 5.6 (95% CI: 1.5–21.0) and LR-of 0.28 (95% CI: 0.11–0.71). Overall, authors' recommendations from all included SR/MA were that clinicians should not use these tests individually because of their poor diagnostic validity and advised combining the results of the tests even though no evidence was presented to support this approach (Hegedus et al., 2007; Hing et al., 2009; Jackson et al., 2003; Meserve et al., 2008; Ryzewicz et al., 2007; Scholten et al., 2001; Smith et al., 2015; Solomon et al., 2001).

3.3.2. Anterior cruciate ligament injuries

Table 4 presents the diagnostic properties of the physical examination tests used for the diagnosis of ACL injuries. Six SR/MA provided data on the diagnosis of ACL injuries (Benjaminse et al., 2006; Jackson et al., 2003; Leblanc et al., 2015; Scholten et al., 2003; Solomon et al., 2001; van Eck et al., 2013) with AMSTAR scores ranging from 2 to 8 out of 11 with a mean of 5.3. Based on the data extracted from *Benjaminse* et al. (Benjaminse et al., 2006), the highest quality meta-analysis with the most studies included in analysis (8/11, n = 28) to also provide 95% CI for LR+/-, the Lachman test demonstrated the highest Sn with 85.0% (95% CI: 83.0–87.0%). For Sn, the point estimates were relatively similar across all included SR/MA (range: 81.0–89.0%). In terms of LR, the Lachman test also reached the highest LR+ and lowest LR-with 10.2 (95% CI: 4.6–22.7) and 0.20 (95% CI: 0.10–0.30) respectively (Benjaminse et al., 2006). The point estimates for the LR + varied across the included SR/MA (range: 4.5–42.0), but was more consistent for the LR- (range: 0.10–0.22). The Pivot Shift test demonstrated the highest Sp with a score of 98.0% (95% CI: 96.0–99.0%) (Benjaminse et al., 2006), but important variations in the point estimates across all included SR/MA was observed (range: 81.0–98.0%). Overall, authors’ recommendations from all included SR/MA were that the Lachman test is of high diagnostic value both to rule in or out an ACL injury while a positive Pivot Shift test can be used to rule in an ACL injury (Benjaminse et al., 2006; Jackson et al., 2003; Scholten et al., 2003; Solomon et al., 2001; van Eck et al., 2013). One review concluded that the anterior drawer test may be used to rule in an ACL injury, but not rule out, and could be used instead of the Lachman in situations where the evaluator has less training performing the test (Benjaminse et al., 2006).

3.3.3. Knee fractures

Table 5 presents the diagnostic properties for the *Ottawa Knee Rule*, a clinical prediction rule used to exclude a knee fracture and avoid unnecessary radiograph of the knee at the emergency (Bachmann et al., 2004). Two reviews providing data on this clinical prediction rule with AMSTAR score of 2 and 7 out of 11 (Bachmann et al., 2004; Jackson et al., 2003). Both reviews reported similar diagnostic properties for the rule. *Bachmann* et al. calculated 98.5% (95% CI: 93.2–100%) for Sn, 48.6% (95% CI: 43.4–51.0%) for Sp and 0.05 (95% CI: 0.02–0.23) for LR- (Bachmann et al., 2004). Overall, the authors' recommendations from both included SR/MA are that this clinical prediction rule, with its low LR-, is considered useful to exclude a fracture if all the rule's criteria are negative; if this is not the case, the clinician cannot rule out a fracture and should order a knee radiograph (Table 5).

3.3.4. Knee osteoarthritis

Table 6 presents the diagnostic properties of one set of

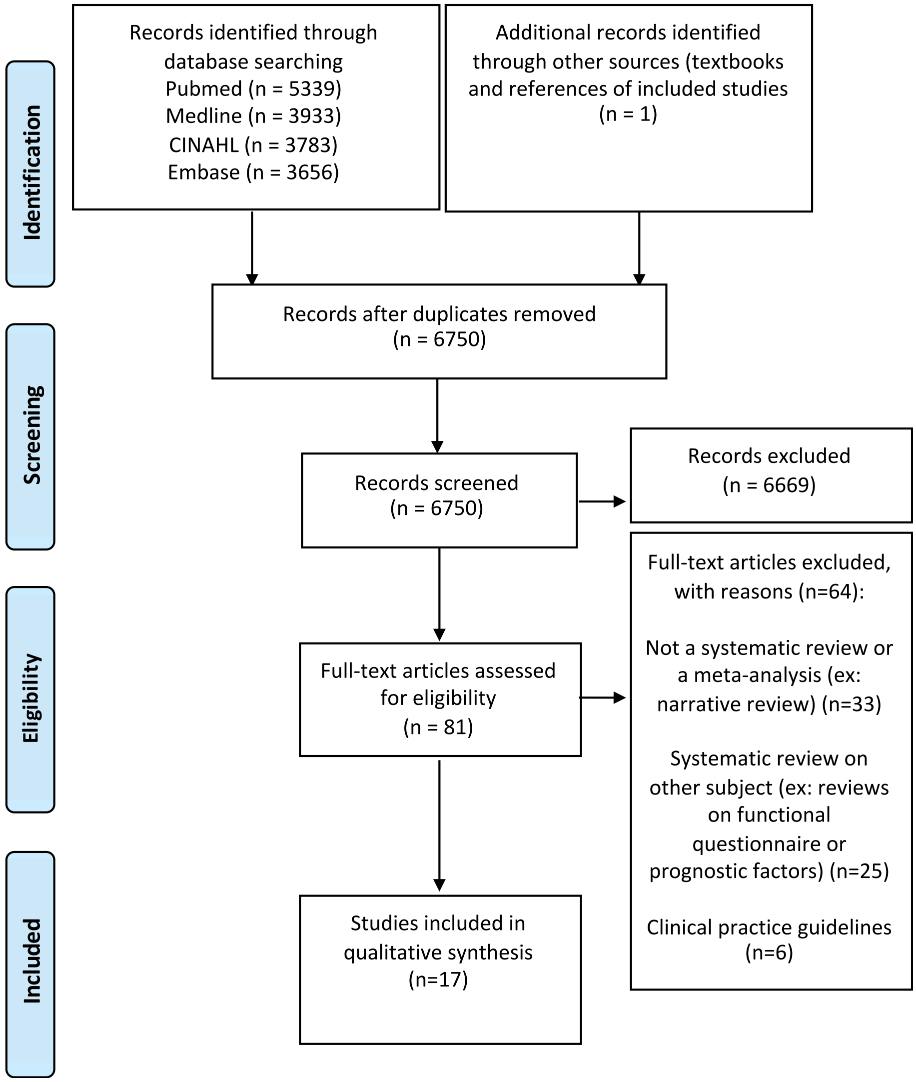


Fig. 1. Bibliographic search flowchart (Liberati et al., 2009; Moher et al., 2009).

diagnostic criteria and one clinical prediction rule for the diagnosis of knee OA. Two reviews were found providing data on such tools with AMSTAR scores of 1 and 2 out of 11 (Jackson et al., 2003; Zhang et al., 2010). The clinical criteria from the *American College of Rheumatology* for the diagnosis of knee OA include: age ≥50 years, stiffness ≤30 min, crepitus, bony tenderness, bony enlargement, and no palpable warmth (Jackson et al., 2003). If at least three criteria are present, the Sn is 95.0% and Sp is 69.0%. If a fourth criterion is present, the Sn drops to 84.0%, but the Sp increases to 89.0%. The study by *Zhang* et al. (EULAR rule) is presented as both a MA with the results of the validation of the rule in two cross-sectional studies (Zhang et al., 2010). For the clinical prediction rule, they reported a post-test probability of 99% for the diagnosis of OA (Kellgren Lawrence ≥2) if six criteria were present based on an estimated prevalence of 12.5% in adults aged ≥ 45 years (Zhang et al., 2010). The criteria include three symptoms: knee pain, morning stiffness for less than 30 min, functional limitations, and three signs: knee crepitus, knee restricted range of motion and bony enlargements (Table 6). In both reviews, the authors concluded that a clinical diagnosis of knee OA can be done with the clinical criteria, but knee radiography remains necessary to assess the radiological grading of OA.

3.3.5. *Patellofemoral pain*

Two reviews evaluated the diagnostic validity of 25 tests for PFP (AMSTAR score: 7 and 6/11) (Cook et al., 2012; Nunes et al., 2013). Both reviews used the same five studies and one included four more for a total of nine studies (Cook et al., 2012; Nunes et al., 2013). Overall, five tests had a LR+≥5: the active instability test (LR+: 249.0), pain during stair climbing (LR+: 11.6), Clarke's sign (LR+: 7.4), pain during prolonged sitting (LR+: 7.5), and patellar tilt (LR+: 5.4, 95%CI:1.4–20.8) (Cook et al., 2012; Nunes et al., 2013). They also reported that pain during squatting demonstrated a LR− = 0.20 (95%CI:0.1–0.4) (Cook et al., 2012; Nunes et al., 2013). Combining tests had a mitigated effect on improving the diagnosis of PFP (Cook et al., 2012; Nunes et al., 2013). However, they acknowledged that the primary diagnostic studies included in their SR reported heterogeneous results with an overall high risk of bias (Cook et al., 2012; Nunes et al., 2013). The authors of the included reviews proposed to view PFP as a diagnosis of exclusion (Cook et al., 2012; Nunes et al., 2013). No individual tests can be recommended at this time to diagnose or exclude a PFP.

3.3.6. *Posterior cruciate ligament injuries*

The SR by *Kopkow* et al. evaluated the diagnostic validity of 11 tests for PCL injuries (AMSTAR score: 7/11) (Kopkow et al., 2013). In

Table 1
Characteristics of the included reviews.

Authors	Study design	Disorders	Gold standard from primary diagnostic studies	Number of studies evaluated	Number of studies selected for pooled analysis	Number of participants included in pooled analyses	AMSTAR Score
(Scholten et al., 2001)	Systematic review and Meta-analysis	Meniscal injuries	Arthroscopy or MRI	N = 13	McMurray's n = 11 Joint line tenderness n = 10 Apley's Test n = 4	N = 2231	6/11
(Hegedus et al., 2007)	Systematic review and meta-analysis	Meniscal injuries	Arthroscopy or MRI	N = 18	McMurray's n = 14 Joint line tenderness n = 14 Apley's test n = 7	N = 2670	8/11
(Ryzewicz et al., 2007)	Systematic Review	Meniscal injuries	Arthroscopy or MRI	N = 32	McMurray's n = 5 Apley's n = 2 Joint line tenderness n = 4	–	2/11
(Meserve et al., 2008)	Systematic review and Meta-analysis	Meniscal injuries	Arthroscopy (MRI accepted in combination)	N = 11	McMurray's n = 8 Joint line tenderness n = 8 Apley's Test n = 3	McMurray's n = 1232 Joint line tenderness n = 1354	6/11
(Hing et al., 2009)	Systematic Review	Meniscal injuries	Arthroscopy or MRI	N = 11	–	–	3/11
(Smith et al., 2015)	Systematic Review and Meta-analysis	Meniscal injuries	Arthroscopy or MRI	N = 9	N = 9	n = 1234	8/11
(Scholten et al., 2003)	Systematic review	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N = 17	Anterior Drawer Test n = 6 Lachman Test n = 6 Pivot Shift Test n = 4	Variable	6/11
(Benjaminse et al., 2006)	Systematic review and meta-analysis	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N = 28	N = 28	Variable	8/11
(van Eck et al., 2013)	Systematic review and meta-analysis	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N = 20	Anterior Drawer Test n = 17 Lachman Test n = 13 Pivot Shift Test n = 12	Anterior Drawer Test n = 1579 Lachman Test n = 934 Pivot Shift Test n = 1192	6/11
(Leblanc et al., 2015)	Systematic review and meta-analysis	Anterior Cruciate Ligament injuries	Arthroscopy or MRI	N = 8	Lachman n = 5 Pivot Shift n = 4	n = 1196	7/11
(Solomon et al., 2001)	Systematic Review	Anterior Cruciate Ligament injuries and Meniscal Injuries	Arthroscopy or MRI	N = 23	ACL n = 15 Meniscal n = 9	Variable	3/11
(Jackson et al., 2003)	Systematic review and meta-analysis	Acute Knee Disorders, Anterior Cruciate Ligament injuries and Meniscal Injuries	Arthroscopy/MRI/radiography/ clinical diagnosis	N = 35	ACL n = 11 Meniscal n = 4	Variable	2/11
(Bachmann et al., 2004)	Systematic review and meta-analysis	Knee fracture with the Ottawa Knee Rule	Radiography or follow-up	N = 11	N = 6	N = 4249	7/11
(Zhang et al., 2010)	Systematic Review, *Meta*-analysis and Delphi consensus	Knee Osteoarthritis	Clinical features and radiographs	N = 313	Variable	Variable	1/11
(Cook et al., 2012)	Systematic review	Patellofemoral pain	Arthroscopy or clinical or imaging (accepted by authors)	N = 9	–	–	7/11
(Nunes et al., 2013)	Systematic review and meta-analysis	Patellofemoral pain	Unreported	N = 5	N = 2	N = 145	6/11
(Kopkow et al., 2013)	Systematic review	Posterior Cruciate Ligament injuries	Arthroscopy or MRI	N = 11	–	–	7/11

MRI: magnetic resonance imaging.

Table 2
Assessment of the methodological quality of the included systematic reviews (AMSTAR).

	Scholten, 2001	Hegedus, 2007	Ryzewicz, 2007	Meserve, 2008	Hing, 2009	Smith, 2015	Scholten, 2003	Benjaminse, 2006	van Eck, 2013	Leblanc, 2015	Solomon, 2001	Jackson, 2003	Bachman, 2004	Zhang, 2010	Cook, 2012	Nunes, 2013	Kopkow, 2013
Item 1	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red	Red
Item 2	Green	Green	Green	Red	Yellow	Yellow	Green	Yellow	Green	Green	Yellow	Yellow	Green	Yellow	Green	Green	Green
Item 3	Green	Green	Yellow	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Yellow	Green	Green	Green
Item 4	Green	Green	Red	Red	Red	Green	Green	Green	Yellow	Red	Red	Red	Red	Red	Green	Yellow	Green
Item 5	Red	Green	Red	Green	Green	Green	Red	Green	Green	Green	Red	Red	Green	Yellow	Green	Green	Green
Item 6	Green	Green	Green	Green	Green	Green	Green	Yellow	Yellow	Green	Green	Red	Green	Yellow	Yellow	Yellow	Green
Item 7	Green	Green	Red	Green	Red	Green	Green	Green	Green	Green	Green	Yellow	Green	Red	Green	Green	Green
Item 8	Green	Green	Red	Green	Yellow	Green	Green	Green	Red	Green	Red	Red	Red	Red	Green	Green	Green
Item 9	Yellow	Green		Yellow		Green	Yellow	Green	Green	Green	Yellow	Green	Green	Yellow		Yellow	
Item 10	Red	Red	Red	Yellow	Yellow	Red	Red	Green	Red	Yellow	Red	Red	Red	Red	Red	Red	Red
Item 11	Red	Red	Red	Green	Red	Green	Red	Green	Green	Red	Green	Green	Green	Green	Green	Green	Red
Score (/11)	**6**	**8**	**2**	**6**	**3**	**8**	**6**	**8**	**6**	**7**	**3**	**2**	**7**	**1**	**7**	**6**	**7**

Item 1: Was an "a priori" design provided?; **Item2:** Was there duplicate study selection and data extraction?; **Item 3:** Was a comprehensive literature search performed?; **Item 4:** Was the status of publication used as an inclusion criterion?; **Item 5:** Was a list of studies provided?; **Item 6:** Were the characteristics of the included studies provided?; **Item 7:** Was the scientific quality of the included studies assessed?; **Item 8:** Was the scientific quality of the included studies used in conclusion?; **Item 9:** Were the methods used to combine the findings appropriate?; **Item 10:** Was the likelihood of publication bias assessed?; **Item 11:** Was the conflict of interest stated?

Green = Yes; Red = No; Yellow = can't say/unclear; Blank = N/A

their review including eleven studies, they reported that the posterior drawer test was the most frequently studied test with a Sn ranging from 22% to 100% (Kopkow et al., 2013). They reported that the included primary study in their SR with the lowest risk of bias (i.e: moderate) showed a LR+ of 50.1 (95% CI: 7.1–351.7) and a LR-of 0.11 (95% CI: 0.03–0.40) for this test (Kopkow et al., 2013). The authors also reported that the quadriceps active test appeared to be the most specific test with a Sp ranging from 96% to 100% based on two included studies from their SR (Kopkow et al., 2013). *Kopkow* et al. concluded that at this time, evidence was insufficient to recommend any physical tests to diagnose or exclude a PCL injury (Kopkow et al., 2013).

3.3.7. History taking and physical examination for the diagnosis of common knee disorders

Table 7 presents the diagnostic properties extracted from reviews on the complete physical examination (including a thorough history and physical tests) for the diagnosis of various knee disorders (Jackson et al., 2003). We found two reviews providing data on such evaluation for four knee disorders with AMSTAR scores of 2 and 3 out of 11 (Jackson et al., 2003; Solomon et al., 2001). For ACL injuries, *Jackson* et al. (Jackson et al., 2003). reported point estimates for LR+ and LR-of 15.0 (95% CI: 5.1–23.0) and 0.27 (95% CI: 0.12–0.42) respectively. For PCL injuries, *Jackson* et al. (Jackson et al., 2003). also reported point estimates for LR+ and LR-of 16.2 (95% CI: 5.2–25.0) and 0.20 (95% CI: 0.13–0.49) respectively. Therefore, a complete physical examination appears valid to diagnose an ACL injury and diagnose or exclude a PCL injury, although the definition of what constitutes a complete physical examination was not provided. Likewise, for the diagnosis of cartilage lesions, the complete physical examination may be considered valid with LR+ of 13.0 (95%CI: 2.7–24.0) and a LR-of 0.51 (95% CI: 0.40–0.62) (Jackson et al., 2003). For the diagnosis of meniscal injuries, *Solomon* et al. reported in their review a LR+: 2.7 (95% CI: 1.4–5.1) and a LR-: 0.40 (95% CI: 0.20–0.70) (Solomon et al., 2001) and concluded that a complete physical examination is not valid to diagnose or exclude a meniscal injury. However, the review by *Jackson* et al. concluded that a complete physical examination might be valid to identify the presence of lateral meniscal injury or to exclude a medial or lateral meniscus injury (Jackson et al., 2003). Overall, studies that reviewed the complete physical examination for various knee disorders concluded that a complete physical examination is probably superior to individual tests but it remains unclear how well this approach may perform (Jackson et al., 2003; Solomon et al., 2001). Based on this evidence, a complete physical examination may be diagnostically superior to individual tests but further research is needed.

4. Discussion

The objective of this study was to systematically review all SR and MA evaluating the diagnostic validity of physical examination tests for common knee disorders performed individually or in combination. This review found a mean AMSTAR score of 5.4 ± 2.4 (95% CI: 4.2–6.6). *Gagnier* et al. recently evaluated the methodological quality of seventy-six orthopaedic SR and MA using the AMSTAR tool and the analysis of this review mirrors their findings (AMSTAR score: 5.4 vs 5.9) (Faggion, 2015; Gagnier & Kellam, 2013). Therefore, in accordance with other published literature, the mean AMSTAR score is at best moderate for the included SR and MA. However, given the wide range of methodological scores, key reviews of high methodological quality were found and the estimate from these reviews may be used to guide clinical decision-making for the diagnosis of common knee disorders. Moreover, conclusions across SR and MA were most often consistent, independently of the year of publication and number of included primary diagnostic studies, reinforcing to a certain extent the strength of the evidence.

Eight SR/MA provided data for the diagnosis of meniscal injuries

Table 3
Description of diagnostic properties for selected tests for meniscal injuries based on results from included reviews.

Tests	Properties	Solomon et al., (2001)	Scholten et al., (2001)	Jackson et al., (2003)	Hegedus et al., (2007)	Ryzewicz et al., (2007)	Meserve et al., (2008)	Hing et al., (2009)	Smith et al., (2015)
McMurray's test	Sn	53.0 ± 15.0%[a]	10.0–66.0%[b]	52.0% [95% CI: 35.0–68.0%]	70.5% [95% CI: 67.4–73.4%]	16.0–67.0%[b]	55.0% [95% CI: 50.0–60.0%]	16.0–88.0%[b]	61.0% [95% CI: 45.0–74.0%]
	Sp	59.0± 36.0%[a]	57.0–98.0%[b]	97.0% [95% CI: 87.0–99.0%]	71.1% [95% CI: 69.3–72.9%]	69.0–98.0%[b]	77.0% [95% CI: 62.0–87.0%]	20.0–98.0%[b]	84.0% [95% CI: 69.0–92.0%]
	LR+	1.3 [95% CI: 0.90–1.7]	1.5–9.5[b]	–	–	–	2.4 [95% CI: 1.3–4.6]	0.82–8.86[b]	3.2 [95% CI: 1.7–5.9]
	LR-	0.80 [95% CI: 0.60–1.1]	0.40–0.90[b]	–	–	–	0.58 [95% CI: 0.46–0.81]	0.24–1.45[b]	0.52 [95% CI: 0.34–0.81]
	DOR	–	–	–	4.5 [95% CI: 3.7–5.4]	–	3.99 [95% CI: 1.04–15.31]	–	–
Apley's maneuver	Sn	–	–	–	60.7% [95% CI: 55.7–65.5%]	16.0–41.0%[b]	22.0% [95% CI: 17.0–28.0%]	–	–
	Sp	–	–	–	70.2% [95% CI: 68–72.4%]	80.0–93.0%[b]	88.0% [95% CI: 72.0–96.0%]	–	–
	DOR	–	–	–	3.4 [95% CI: 2.6–4.4]	–	2.20 [95% CI: 0.27–17.66]	–	–
Joint line tenderness	Sn	79.0 ± 4.0%[a]	28.0–95.0%[b]	76.0% [95% CI: 65.0–87.0%]	63.3% [95% CI: 60.9–65.7%]	67.0–97.0%[b]	76.0% [95% CI: 73.0–80.0%]	–	83.0% [95% CI: 73.0–90.0%]
	Sp	15.0± 22.0%[a]	13.0–95.0%[b]	29.0% [95% CI: 10.0–46.0%]	77.4% [95% CI: 75.6–79.1%]	29.4–87.0%[b]	77.0% [95% CI: 64.0–87.0%]	–	83% [95% CI: 61.0–94.0%]
	LR+	0.90 [95% CI: 0.80–1.0]	0.80–14.9[b]	–	–	–	3.3 [95% CI: 1.6–6.2]	–	4.0 [95% CI: 2.1–7.5]
	LR-	1.1 [95% CI: 1.0–1.3]	0.20–2.1[b]	–	–	–	0.31 [95% CI: 0.23–0.42]	–	0.23 [95% CI: 0.12–0.44]
	DOR	–	–	–	4.5 [95% CI: 3.8–5.4]	–	10.98 [95% CI: 3.02–39.95]	–	–
Number of primary diagnostic studies included in analysis		N = 9	N = 13	N = 4	N = 18	N = 11	N = 11	N = 11	N = 9
AMSTAR Score		3/11	6/11	2/11	8/11	2/11	6/11	3/11	8/11

Sn: sensitivity, Sp: specificity, LR: likelihood ratio, DOR: diagnostic odds ratio.
[a] Data presented as mean ± SD (standard deviation), calculated by authors.
[b] Indicates that the authors could not pool the data and did not calculate a mean value with standard deviation. Therefore, we presented the range of values based on the data presented in the article.

Table 4
Description of diagnostic properties for physical examination tests for Anterior Cruciate Ligament (ACL) injuries based on results from included reviews.

Tests	Properties	Solomon et al., (2001)	Jackson et al., (2003)	Scholten et al., (2003)	Benjaminse et al., (2006)	van Eck et al., (2013) [c]	Leblanc et al., (2015)
Lachman test	Sn	84.0 ± 15.0%[a]	87.0% [95% CI: 76.0–98.0%]	86.0% [95% CI: 76.0–92.0%]	85.0% [95% CI: 83.0–87.0%]	81.0%[c]	89.0% [95% CI: 76.0–98.0%]
	Sp	100%	93.0% [95% CI: 89.0–96.0%]	91.0% [95% CI: 79.0–96.0%]	94.0% [95% CI: 92.0–95.0%]	81.0% [c]	–
	LR+	42.0 [95% CI: 2.7–651.0]	–	2.0–102.1[b]	10.2 [95% CI: 4.6–22.7]	4.5[c]	–
	LR-	0.10 [95% CI: 0.00–0.40]	–	0.10–0.40[b]	0.20 [95% CI: 0.10–0.30]	0.22[c]	–
	DOR	–	–	–	70.0 [95% CI: 23.0–206.0]	–	–
Pivot shift test	Sn	38.0± 28.0%[a]	61.0% [95% CI: 40.0–82.0%]	18.0–48.0%[b]	24.0% [95% CI: 21.0–27.0%]	28.0% [c]	79% [95% CI: 63.0–91.0%]
	Sp	–	97.0% [95% CI: 93.0–99.0%]	97.0–99.0%[b]	98.0% [95% CI: 96.0–99.0%]	81.0% [c]	–
	LR+	–	–	8.2–26.9[b]	8.5 [95% CI: 4.7–15.5]	5.35[c]	–
	LR-	–	–	0.50–0.80[b]	0.90 [95% CI: 0.80–1.0]	0.84[c]	–
	DOR	–	–	–	12.0 [95% CI: 5.0–31.0]	–	–
Anterior drawer test	Sn	62.0 ± 23.0%[a]	48.0% [95% CI: 38.0–59.0%]	62.0% [95% CI: 42.0–78.0%]	55.0% [95% CI: 52.0–58.0%]	38.0% [c]	–
	Sp	67.0± 42.0%[a]	87.0% [95% CI: 83.0–91.0%]	88.0% [95% CI: 83.0–92.0%]	92.0% [95% CI: 90.0–94.0%]	81.0% [c]	–
	LR+	3.8 [95% CI: 0.70–22.0]	–	1.7–87.9[b]	7.3 [95% CI: 3.5–15.2]	4.52[c]	–
	LR-	0.30 [95% CI: 0.05–1.50]	–	0.10–0.80[b]	0.50 [95% CI: 0.40–0.60]	0.67[c]	–
	DOR	–	–	–	21.0 [95% CI: 8.0–53.0]	–	–
Number of primary diagnostic studies included in analysis		N = 15	N = 11	N = 17	N = 28	N = 20	N = 8
AMSTAR Score		3/11	2/11	6/11	8/11	6/11	7/11

Sn: sensitivity, Sp: specificity, LR: likelihood ratio, DOR: diagnostic odds ratio.
[a] Data presented as mean ± SD (standard deviation), calculated by authors.
[b] Authors did not pool data. Ranges of values are therefore presented.
[c] 95% CI not presented in review.

(Hegedus et al., 2007; Hing et al., 2009; Jackson et al., 2003; Meserve et al., 2008; Ryzewicz et al., 2007; Scholten et al., 2001; Smith et al., 2015; Solomon et al., 2001), two SR provided data for the diagnosis of PFP (Cook et al., 2012) and one for the diagnosis of PCL injuries (Kopkow et al., 2013). Overall, the authors' conclusions regarding the evidence supporting the use of physical tests performed individually to diagnose one of these common knee disorders is inconclusive. It appears that none of the physical tests studied (McMurray's, Apley's manoeuver, joint line tenderness and Thessaly test) are able to make a valid diagnosis of a meniscal injury when performed individually, even in high-quality reviews. Likewise, based on moderate quality reviews when performed individually, relevant physical tests are unable to diagnose or exclude a PFP or a PCL injury.

Table 5
Description of diagnostic properties for selected tests for knee fractures based on results from included reviews.

Test	Properties	Jackson et al., (2003)	Bachmann et al., (2004)
Ottawa Knee Rule[a]	Sn	100.0% [95% CI: 94.0–100.0%]	98.5% [95% CI: 93.2–100%]
	Sp	49.0% [95% CI: 46.0–52.0%]	48.6% [95% CI: 43.4–51.0%]
	LR +	1.96 [95% CI: 1.92–1.99]	–
	LR -	0.11 [95% CI: 0.06–0.18]	0.05 [95% CI: 0.02–0.23]
Number of primary diagnostic studies included in analysis		N = 2	N = 6
AMSTAR Score		2/11	7/11

Sn: sensitivity, Sp: specificity, LR: likelihood ratio.
[a] Ottawa Knee Rule criteria (Bachmann et al., 2004; Jackson et al., 2003): A knee radiograph examination is required only for patients who have at least ONE of the followings: 1) aged 55 years or older, 2) isolated tenderness of patella (no bone tenderness of knee other than patella), 3) tenderness at the head of fibula, 4) inability to flex knee to 90°, or 5) inability to bear weight both immediately and in the emergency department for 4 steps (unable to transfer weight twice onto each lower limb regardless of limping).

Table 6
Description of diagnostic properties for selected tests for osteoarthritis based on results from included reviews.

Test	Properties	Solomon et al., (2001)	Zhang et al., (2010)
American College of Rhumatology Criteria[b]	Sn (at least 3 criteria)	95.0%[a]	–
	Sp (at least 3 criteria)	69.0%[a]	–
	Se (4 criteria present)	84.0%[a]	–
	Sp (4 criteria present)	89.0%[a]	–
EULAR Rule[c]	Post test probability	–	99.0%[a]
Number of primary diagnostic studies included in analysis		Unclear	Unclear
AMSTAR score		2/11	1/11

Sn: sensitivity, Sp: specificity.
[a] Data presented directly from authors.
[b] The clinical criteria from the ACR for the diagnosis of osteoarthritis (Jackson et al., 2003) include: age older than 50 years, stiffness for less than 30 min, crepitus, bony tenderness, bony enlargement, and no palpable warmth. Based on a prevalence of 34% in primary care, the presence of at least 3 criteria increases the probability of having osteoarthritis to 62%.
[c] EULAR's clinical prediction rule criteria (Zhang et al., 2010) for the diagnosis of osteoarthritis include: 3 symptoms: knee pain, morning stiffness for less than 30 min, functional limitations, and 3 signs: crepitus, restricted movement and bony enlargement. Based on a background prevalence of 12.5% in adults age more or equal to 45 years, the post-test probability of having radiographic knee OA ($KL \geq 2$) increase with the number of criteria present to 99% when all six criteria are present.

Six SR/MA provided data for the diagnosis of ACL injuries (Benjaminse et al., 2006; Jackson et al., 2003; Leblanc et al., 2015; Scholten et al., 2003; Solomon et al., 2001; van Eck et al., 2013). The evidence suggests that out of the three tests that were investigated (Lachman, anterior drawer test and the pivot shift), the Lachman appears to be valid when individually performed to diagnose or exclude an ACL injury and the Pivot Shift may be used to diagnose an ACL injury because of its high specificity. Of note, one SR included in the present analysis concluded that the use of the pivot shift for patients under anaesthesia reached higher specificity than the Lachman (van Eck et al., 2013). Clinically, the evidence indicates that clinicians can rely on the Lachman test to rule in or out an ACL injury. It should be noted that some authors expressed concerns about the use of the anterior drawer test in clinics because of its lower validity compared to the Lachman test even if this test reaches an adequate validity (Jackson et al., 2003).

Two reviews provided data for the screening of knee fracture using the *Ottawa Knee Rule* (Bachmann et al., 2004; Jackson et al., 2003). The evidence indicates that the *Ottawa Knee Rule* is valid to exclude knee fractures (Stiell et al., 1997) and has been used extensively in clinical practice since its original development two decades ago. The implementation of this tool in clinical practice, namely in emergency departments, has been shown to decrease radiograph use, to lower direct healthcare costs and to diminish time spent by patients in the emergency department, without compromising patients' safety and quality of care. (Perry & Stiell, 2006; Stiell & Bennett, 2007). Clinicians can therefore rely on the rule to avoid unnecessary radiographic evaluation in this setting.

Two reviews provided data for the diagnosis of OA using diagnostic criteria and a clinical prediction rule (Jackson et al., 2003; Zhang et al., 2010). The *American College of Rheumatology* has issued criteria for the diagnosis of osteoarthritis of the knee without the use of radiographs, but only the sensitivity and specificity are reported, a fact that limits clinical interpretation (Jackson et al., 2003). Moreover, initial development of the criteria mainly classify patients from having OA or rheumatoid arthritis (Altman et al., 1986) and there remains conjecture in the literature about the use of the criteria in primary care (Peat et al., 2006). The recent EULAR prediction rule provides LR+ and LR- and therefore enables the calculation of post-test probability (Zhang et al., 2010). The post-test probability of this rule is very high if all six symptoms and signs are present. However, to our knowledge, the diagnostic capabilities of this combination when fewer than six criteria are positive has not been presented, which limits its usefulness for clinicians. Clinically, the evidence indicates that clinicians may use the proposed criteria to diagnose clinical knee OA without the use of radiographs as proposed by both medical associations. However, radiographs remain necessary for radiological grading as can be used to plan a consultation with an orthopaedic surgeon. More evidence is needed to better define the clinical diagnosis of knee OA in the general population (Jackson et al., 2003; Peat et al., 2006).

The authors from all the included SR/MA argued that the majority of primary diagnostic studies were of low to medium methodological quality. Commonly cited poor methodological aspects included a low number of patients and/or of cases, a retrospective study design, the lack of blinding, high risk of spectrum bias and inadequate description of index tests and reference standards (Hegedus et al., 2007; Jackson et al., 2003). These biases have been shown to inflate the diagnostic validity of the index test under study (Whiting et al., 2013). Moreover, only two reviews reported the reliability of clinical tests (Cook et al., 2012; Zhang et al., 2010). Reliability is an important and commonly overlooked diagnostic quality that ultimately influences the validity of the test and its usefulness for clinicians (Desmeules et al., 2013). More precise discussion on this subject is beyond the scope of this review, but evaluating the available evidence on the reliability of knee physical tests is warranted.

The evidence presented above demonstrates that the diagnosis of common knee disorders cannot be made by relying exclusively on one physical test, with exception of ACL injuries. The presented evidence can inform clinicians of the validity of physical tests that may be useful to diagnose or exclude a disorder. Interestingly, our review presents evidence, although limited and in low quality reviews, that a complete physical examination may be diagnostically superior to individual tests. However, in the present review, the definition of the history elements and physical tests included in a complete physical examination was notably lacking. Indeed, we do know that clinicians rely on combinations of history elements and physical tests to make a diagnosis when patients present with a knee problem. It has been reported that clinicians with extensive training with musculoskeletal disorders are able to diagnose knee disorders with high confidence using a complete physical examination (Daker-White et al., 1999; Desmeules et al., 2013; Moore et al., 2005). Our review also shows that when clinical prediction rules are developed they may become useful, valid tools for the diagnosis of common knee disorders such as OA and fractures. Of interest, several clinical prediction rules have been developed in recent publications not presented in the included reviews (Cook et al., 2010; Kastelein et al., 2008, 2009; Wagemakers et al., 2008; Wagemakers et al., 2010) and demonstrated that the combination of specific history elements and physical tests have improved the diagnostic validity compared to individual tests for specific knee disorders. We believe that more methodologically sound diagnostic studies are needed and should focus on the evaluation of the

Table 7
Description of diagnostic properties of the history taking and physical examination for selected knee disorders based on results from included reviews.

Disorders	Properties	Solomon et al., (2001)	Jackson et al., (2003)
Anterior cruciate ligament[b]	Sn	82.0%[a]	74.0% [95% CI: 60.0–88.0%]
	Sp	94.0%[a]	95.0% [95% CI: 92.0–98.0%]
	LR+	25.0 [95% CI: 2.1–306.0]	15.0 [95% CI: 5.1–23.0]
	LR-	0.04 [95% CI: 0.01–0.48]	0.27 [95% CI: 0.12–0.42]
Posterior cruciate ligament[b]	Sn	91.0%[a]	81.0 [95% CI: 63.0–98.0%]
	Sp	98.0%[a]	95.0 [95% CI: 81.0%–100.0%]
	LR+	21.0 [95% CI: 2.1–205.0]	16.2 [95% CI: 5.2–25.0]
	LR-	0.05 [95% CI: 0.01–0.50]	0.20 [95% CI: 0.13–0.49]
Cartilage lesions[b]	Sn	–	0.51 [95% CI: 0.37–0.65]
	Sp	–	0.96 [95% CI: 0.91–1.0]
	LR+	–	13.0 [95% CI: 2.7–24.0]
	LR-	–	0.51 [95% CI: 0.40–0.62]
Any meniscus[b]	Sn	77.0± 7.0%[a]	–
	Sp	91.0± 3.0%[a]	–
	LR+	2.7 [95% CI: 1.4–5.1]	–
	LR-	0.40 [95% CI: 0.20–0.70]	–
Medial meniscus[b]	Sn	–	86.0% [95% CI: 79.0–92.0%]
	Sp	–	72.0% [95% CI: 61.0–83.0%]
	LR+	–	3.1 [95% CI: 0.54–5.7]
	LR-	–	0.19 [95% CI: 0.11–0.77]
Lateral meniscus[b]	Sn	–	88.0% [95% CI: 77.0–99.0%]
	Sp	–	92.0% [95% CI: 89.0–95.0%]
	LR+	–	11.0 [95% CI: 1.8–20.2]
	LR-	–	0.13 [95% CI: 0.00–0.25]
Number of primary diagnostic studies included in analysis		N = 3	N = 35
AMSTAR Score		N = 3/11	2/11

Sn: sensitivity, Sp: specificity, LR: likelihood ratio.
[a] Data presented as mean ± SD (standard deviation), calculated or presented directly from the authors.
[b] The examination tests used in the primary diagnostic studies on the complete physical examination varies greatly and are often partially or not reported (Jackson et al., 2003).

clinical prediction rules that incorporate well-defined and specific history elements and physical tests.

5. Strengths and limitations of the present review

This SR has the advantage to provide an overview of the diagnostic validity for common knee disorders. Every included SR/MA was methodologically assessed using the AMSTAR. A limitation of this SR was the difficulty to combine the point estimates of SR and MA. We chose to present the range of the point estimates across included SR/MA to obtain an overall qualitative appraisal of the evidence. Moreover, the analysis of the range of the point estimates across the included SR/MA provided an opportunity to show the frequent wide range and heterogeneity of the evidence.

Future SRs could be improved by following the established reporting guidelines (Liberati et al., 2009; Moher et al., 2009). In particular, presentation of a priori design, meta-analysis, assessment of publication bias and presentation of conflict of interests were rarely present. Improving the quality of primary diagnostic studies would enable the meta-analysis and more precise estimate of the true diagnostic value of certain tests. Lastly, the aim of this SR was to synthesize all SR/MA and therefore some new and recent primary diagnostic studies evaluating new physical examination tests or clinical prediction rules were not part of our analyses. Nonetheless, we believe that our review provides useful recommendations to clinicians and researchers working with patients suffering from common knee disorders.

6. Conclusions

Many SR and MA are of low to moderate quality, which warrants caution from clinicians when reading these reviews for clinical guidance. However, a few methodologically sound reviews provide high-quality evidence for ACL and meniscal injuries. The evidence suggests that clinicians may diagnose or exclude an ACL injury with the Lachman test, exclude a knee fracture using the *Ottawa Knee Rule* and make a diagnosis of knee OA based on the results of the *American College of Rhumatology and EULAR* rules. For other knee disorders (meniscal injury, PFP, PCL injury and others), the available evidence does not demonstrate that tests used individually are diagnostically valid. Globally, very few clinical tests, when performed individually, can diagnose or exclude a knee disorder. Based on limited and low quality evidence, the combination of history elements and physical tests may be more diagnostically valid. In the context of increasing healthcare costs, the development of clinical prediction rules comprising history elements and physical examination tests from methodologically sound diagnostic studies are necessary to further advance the diagnosis of knee disorders.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This project received funding from Canadian Institute of Health Research (CIHR) 349856. SD obtained doctoral support from Fond de Recherche en Santé du Québec (FRQS) 28767.

Appendix 1. detailed search strategy with keywords and descriptors.

#	Search details	Results
Pubmed		
1	(((((Reliab* OR reproduc* OR validity OR accuracy OR variability OR “predictive value”)) OR (interobserver* OR intertester* OR interrater* OR interexamin* OR intraobserver* OR intratester* OR intrarater* OR intraexamin* OR “observer variation” [Mesh])) OR ((inter OR intra) and (rater* OR examin* OR tester* OR observer*)))) and ((((patholog* OR lesion* OR ruptur$ OR torn OR tear* OR trauma OR traumas OR effusion* OR instabilit* OR laxity OR injur* OR disorder* OR syndrome OR pain OR osteoarthr* OR alignment)) and (knee* OR “anterior cruciate ligament*” OR “posterior cruciate ligament*” OR “medial collateral ligament*” OR menisc* OR patell* OR “knee” [Mesh] OR “knee joint” [Mesh] OR “ligaments, articular” [Mesh] OR “patella” [Mesh])) and (“physical examination” OR “physical examination” [Mesh] OR “clinical assessment” OR test OR tests OR manual* OR manoeuv* OR maneuv* OR palpation OR examiner*))	5345
Medline		
1	(patholog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instabilit$ or laxity or injur$ or disorder$ or syndrome or pain or osteoarthr$ or alignment). mp.	4 840 034
2	(knee$ or “anterior cruciate ligament$” or “posterior cruciate ligament*” or “medial collateral ligament*” or menisc$ or patell$). mp. or knee/ or exp knee joint/or exp ligaments, articular/or patella/	144 958
3	(“physical examination” or “clinical assessment” or examiner$ or test$ or manual$ or man?euv$ or palpation). mp. or exp physical examination/	4 110 216
4	(reliab$ or reproduc$ or validity or accuracy or variability). mp.	1 312 329
5	((inter or intra) adj (rater$ or examin$ or tester$ or observer$)). mp.	12 210
6	(interobserver$ or intertester$ or interrater$ or interexamin$ or intraobserver$ or intratester$ or intrarater$ or intraexamin$). mp. or observer variation/	46 037
7	4 or 5 or 6	1 325 926
8	1 and 2 and 3 and 7	3935
CINAHL		
1	patholog* or lesion* or ruptur* or torn or tear* or trauma$ or effusion* or instabilit* or laxity or injur* or disorder* or syndrome or pain or osteoarthr* or alignment	965 537
2	knee or knees or “anterior cruciate ligament*” or “posterior cruciate ligament*” or “medial collateral ligament*” or menisc$ or patell* or MH “Knee” or MH “Knee Joint+” or MH “Patella” OR MH “Patellar Ligament” or MH “Anterior Cruciate Ligament” or MH “Medial Collateral Ligament, knee” or MH “Posterior Cruciate Ligament”	44 735
3	“physical examination” or “clinical assessment” or examiner* or test* or manual* or maneuv* or manoeuv* or palpation or MH “Physical Examination”	700 651
4	reliab* or reproduc* or validity or accuracy or variability	201 057
5	MH “Reliability” or MH “Reliability and Validity” or MH “Interrater reliability” or MH “Intrarater reliability” or “Test-Retest Reliability”	48 676
6	((inter or intra) N (rater* or examin* or tester* or observer*	292 870
7	interobserver* or intertester* or interrater* or interexamin* or intraobserver* or intratester* or intrarater* or intraexamin*	24 654
8	S4 OR S5 OR S6 OR S7	452 480
9	S1 AND S2 AND S3 AND S8	3783
EMBASE		
1	(patholog$ or lesion$ or ruptur$ or torn or tear$ or trauma$ or effusion$ or instabilit$ or laxity or injur$ or disorder$ or syndrome or pain or osteoarthr$ or alignment). mp.	6 587 026
2	(knee or knees or “ anterior cruciate ligament$ ” or “ posterior cruciate ligament$ ” or “ medial collateral ligament$ ” or menisc$ or patell$). mp. or knee/or exp knee ligament/or patella ligament/or patella/or knee meniscus/	176 966
3	(“ physical examination ” or “ clinical assessment ” or examiner$ or test$ or manual$ or man?euv$ or palpation). tw, kw. Or exp physical examination/	3 319 371
4	(reliab$ or reproduc$ or validity or accuracy or variability). mp.	1 846 828
5	((inter or intra) adj (rater$ or examin$ or tester$ or observer$)). mp.	17 804
6	(interobserver$ or intertester$ or interrater$ or interexamin$ or intraobserver$ or intratester$ or intrarater$ or intraexamin$). mp. or observer variation/	41 503
7	intrarater reliability/or interrater reliability/or test retest reliability/or reliability/	112 517
8	4 or 5 or 6 or 7	1 856 756
9	1 and 2 and 3 and 8	3660

References

Altman, R., Asch, E., Bloch, D., Bole, G., Borenstein, D., Brandt, K., Christy, W., Cooke, T. D., Greenwald, R., Hochberg, M., et al. (1986). Development of criteria

for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. *Arthritis Rheum, 29*, 1039–1049.
Andersen, C. A., Clarsen, B., Johansen, T. V., & Engebretsen, L. (2013). High prevalence of overuse injury among iron-distance triathletes. *Br J Sports Med, 47*, 857–861.
Bachmann, L. M., Haberzeth, S., Steurer, J., & ter Riet, G. (2004). The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. *Ann Intern Med, 140*, 121–124.
Baker, P., Reading, I., Cooper, C., & Coggon, D. (2003). Knee disorders in the general population and their relation to occupation. *Occup Environ Med, 60*, 794–797.
Barber Foss, K. D., Myer, G. D., Chen, S. S., & Hewett, T. E. (2012). Expected prevalence from the differential diagnosis of anterior knee pain in adolescent female athletes during preparticipation screening. *J Athl Train, 47*, 519–524.
Benjaminse, A., Gokeler, A., & van der Schans, C. P. (2006). Clinical diagnosis of an anterior cruciate ligament rupture: A meta-analysis. *The Journal of Orthopaedic & Sports Physical Therapy, 36*(5), 267–288.
Bharam, S., Vrahas, M. S., & Fu, F. H. (2002). Knee fractures in the athlete. *Orthopedic Clinics of North America, 33*(3), 565–574.
Cook, C., Hegedus, E., Hawkins, R., Scovell, F., & Wyland, D. (2010). Diagnostic accuracy and association to disability of clinical test findings associated with patellofemoral pain syndrome. *Physiother Can, 62*, 17–24.
Cook, C., Mabry, L., Reiman, M. P., & Hegedus, E. J. (2012). Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. *Physiotherapy, 98*, 93–100.
Cook, C., & Hegedus, E. (2011). *Orthopedic physical examination Tests: An evidence-based approach* (2nd ed.). Prentice Hall.
Daker-White, G., Carr, A. J., Harvey, I., Woolhead, G., Bannister, G., Nelson, I., & Kammerling, M. (1999). A randomised controlled trial. Shifting boundaries of doctors and physiotherapists in orthopaedic outpatient departments. *J Epidemiol Community Health, 53*, 643–650.
Desmeules, F., Toliopoulos, P., Roy, J. S., Woodhouse, L. J., Lacelle, M., Leroux, M., Girard, S., Feldman, D. E., & Fernandes, J. C. (2013). Validation of an advanced practice physiotherapy model of care in an orthopaedic outpatient clinic. *BMC Musculoskelet Disord, 14*, 162.
van Eck, C. F., van den Bekerom, M. P., Fu, F. H., Poolman, R. W., & Kerkhoffs, G. M. (2013). Methods to diagnose acute anterior cruciate ligament rupture: a meta-analysis of physical examinations with and without anaesthesia. *Knee Surg Sports Traumatol Arthrosc, 21*, 1895–1903.
Faggion, C. M., Jr. (2015). Critical appraisal of AMSTAR: Challenges, limitations, and potential solutions from the perspective of an assessor. *BMC Medical Research Methodology, 15*, 63.
Filbay, S. R., Culvenor, A. G., Ackerman, I. N., Russell, T. G., & Crossley, K. M. (2015). Quality of life in anterior cruciate ligament-deficient individuals: a systematic review and meta-analysis. *Br J Sports Med, 49*, 1033–1041.
Gagnier, J. J., & Kellam, P. J. (2013). Reporting and methodological quality of systematic reviews in the orthopaedic literature. *The Journal of Bone & Joint Surgery*, (11), 95. p. e771-7.
Granan, L. P., Inacio, M. C., Maletis, G. B., Funahashi, T. T., & Engebretsen, L. (2013). Sport-Specific Injury Pattern Recorded During Anterior Cruciate Ligament Reconstruction. *Am J Sports Med, 41*(12), 2814–2818.
Hegedus, E. J., Cook, C., Hasselblad, V., Goode, A., & McCrory, D. C. (2007). Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. *J Orthop Sports Phys Ther, 37*, 541–550.
Hegedus, E. J., Goode, A. P., Cook, C. E., Michener, L., Myer, C. A., Myer, D. M., & Wright, A. A. (2012). Which physical examination tests provide clinicians with the most value when examining the shoulder? Update of a systematic review with meta-analysis of individual tests. *Br J Sports Med, 46*, 964–978.
Hing, W., White, S., Reid, D., & Marshall, R. (2009). Validity of the McMurray's Test and Modified Versions of the Test: A Systematic Literature Review. *J Man Manip Ther, 17*, 22–35.
Jackson, J. L., O'Malley, P. G., & Kroenke, K. (2003). Evaluation of acute knee pain in primary care. *Annals of Internal Medicine, 139*(7), 575–588.
Jaeschke, R., Guyatt, G. H., & Sackett, D. L. (1994). 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, 271*(9), 703–707.
Jette, A. M., & Delitto, A. (1997). Physical therapy treatment choices for musculoskeletal impairments. *Physical Therapy, 77*(2), 145–154.
Kastelein, M., Wagemakers, H. P., Luijsterburg, P. A., Verhaar, J. A., Koes, B. W., & Bierma-Zeinstra, S. M. (2008). Assessing medial collateral ligament knee lesions in general practice. *Am J Med, 121*, 982–988. e2.
Kastelein, M., Luijsterburg, P. A., Wagemakers, H. P., Bansraj, S. C., Berger, M. Y., Koes, B. W., & Bierma-Zeinstra, S. M. (2009). Diagnostic value of history taking and physical examination to assess effusion of the knee in traumatic knee patients in general practice. *Arch Phys Med Rehabil, 90*, 82–86.
Kopkow, C., Freiberg, A., Kirschner, S., Seidler, A., & Schmitt, J. (2013). Physical examination tests for the diagnosis of posterior cruciate ligament rupture: a systematic review. *J Orthop Sports Phys Ther, 43*, 804–813.
Laslett, L. L., Quinn, S. J., Winzenberg, T. M., Sanderson, K., Cicuttini, F., & Jones, G. (2012). A prospective study of the impact of musculoskeletal pain and radiographic osteoarthritis on health related quality of life in community dwelling older people. *BMC Musculoskelet Disord, 13*, 168.
Leblanc, M. C., Kowalczuk, M., Andruszkiewicz, N., Simunovic, N., Farrokhyar, F., Turnbull, T. L., Debski, R. E., & Ayeni, O. R. (2015). Diagnostic accuracy of physical examination for anterior knee instability: a systematic review. *The Knee Surgery, Sports Traumatology, Arthroscopy, 23*(10), 2805–2813.
Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med, 6*, e1000100.
Magee, D. J. (2014). *Orthopedic physical assessment, 6th edition*. Saunders, 1184.
Meserve, B. B., Cleland, J. A., & Boucher, T. R. (2008). A meta-analysis examining clinical test utilities for assessing meniscal injury. *Clinical Rehabilitation, 22*(2), 143–161.
Mody, G. M., & Brooks, P. M. (2012). Improving musculoskeletal health: Global issues. *Best Practice & Research Clinical Rheumatology, 26*(2), 237–249.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med, 6*, e1000097.
Moore, J. H., Goss, D. L., Baxter, R. E., DeBerardino, T. M., Mansfield, L. T., Fellows, D. W., & Taylor, D. C. (2005). Clinical diagnostic accuracy and magnetic resonance imaging of patients referred by physical therapists, orthopaedic surgeons, and nonorthopaedic providers. *J Orthop Sports Phys Ther, 35*, 67–71.
Nunes, G. S., Stapait, E. L., Kirsten, M. H., de Noronha, M., & Santos, G. M. (2013). Clinical test for diagnosis of patellofemoral pain syndrome: Systematic review with meta-analysis. *Phys Ther Sport, 14*, 54–59.
Peat, G., Thomas, E., Duncan, R., Wood, L., Hay, E., & Croft, P. (2006). Clinical classification criteria for knee osteoarthritis: performance in the general population and primary care. *Ann Rheum Dis, 65*, 1363–1367.
Perry, J. J., & Stiell, I. G. (2006). Impact of clinical decision rules on clinical care of traumatic injuries to the foot and ankle, knee, cervical spine, and head. *Injury, 37*(12), 1157–1165.
Rahman, L. A., Adie, S., Naylor, J. M., Mittal, R., So, S., & Harris, I. A. (2013). A systematic review of the diagnostic performance of orthopedic physical examination tests of the hip. *BMC Musculoskelet Disord, 14*, 257.
Reiman, M. P., Goode, A. P., Hegedus, E. J., Cook, C. E., & Wright, A. A. (2013). Diagnostic accuracy of clinical tests of the hip: a systematic review with meta-analysis. *Br J Sports Med, 47*, 893–902.
Ryzewicz, M., Peterson, B., Siparsky, P. N., & Bartz, R. L. (2007). The diagnosis of meniscus tears: the role of MRI and clinical examination. *Clin Orthop Relat Res, 455*, 123–133.
Scholten, R. J., Deville, W. L., Opstelten, W., Bijl, D., van der Plas, C. G., & Bouter, L. M. (2001). The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis. *J Fam Pract, 50*, 938–944.
Scholten, R. J., Opstelten, W., van der Plas, C. G., Bijl, D., Deville, W. L., & Bouter, L. M. (2003). Accuracy of physical diagnostic tests for assessing ruptures of the anterior cruciate ligament: a meta-analysis. *J Fam Pract, 52*, 689–694.
Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007a). Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol, 7*, 10.
Shea, B. J., Bouter, L. M., Peterson, J., Boers, M., Andersson, N., Ortiz, Z., Ramsay, T., Bai, A., Shukla, V. K., & Grimshaw, J. M. (2007b). External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One, 2*, e1350.
Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., Henry, D. A., & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol, 62*, 1013–1020.
Skou, S. T., Thomsen, H., & Simonsen, O. H. (2014). The value of routine radiography in patients with knee osteoarthritis consulting primary health care: A study of agreement. *European Journal of General Practice, 20*(1), 10–16.
Smith, B. E., Thacker, D., Crewesmith, A., & Hall, M. (2015). Special tests for assessing meniscal tears within the knee: a systematic review and meta-analysis. *Evid Based Med, 20*, 88–97.
Snoeker, B. A., Bakker, E. W., Kegel, C. A., & Lucas, C. (2013). Risk factors for meniscal tears: A systematic review including meta-analysis. *The Journal of Orthopaedic & Sports Physical Therapy, 43*(6), 352–367.
Solomon, D. H., Simel, D. L., Bates, D. W., Katz, J. N., & Schaffer, J. L. (2001). The rational clinical examination. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. *JAMA, 286*, 1610–1620.
Stein, L. A. (2005). Making the best use of radiological resources in Canada. *Healthc Papers, 6*(1), 18–23.
Stiell, I. G., & Bennett, C. (2007). Implementation of clinical decision rules in the emergency department. *Academic Emergency Medicine, 14*(11), 955–959.
Stiell, I. G., Wells, G. A., Hoag, R. H., Sivilotti, M. L., Cacciotti, T. F., Verbeek, P. R., Greenway, K. T., McDowell, I., Cwinn, A. A., Greenberg, G. H., Nichol, G., & Michael, J. A. (1997). Implementation of the Ottawa Knee Rule for the use of radiography in acute knee injuries. *JAMA, 278*, 2075–2079.
Tveit, M., Rosengren, B. E., Nilsson, J. A., & Karlsson, M. K. (2012). Former male elite athletes have a higher prevalence of osteoarthritis and arthroplasty in the hip and knee than expected. *Am J Sports Med, 40*, 527–533.
Wagemakers, H. P., Heintjes, E. M., Boks, S. S., Berger, M. Y., Verhaar, J. A., Koes, B. W., & Bierma-Zeinstra, S. M. (2008). Diagnostic value of history-taking and physical examination for assessing meniscal tears of the knee in general practice. *Clin J Sport Med, 18*, 24–30.
Wagemakers, H. P., Luijsterburg, P. A., Boks, S. S., Heintjes, E. M., Berger, M. Y., Verhaar, J. A., Koes, B. W., & Bierma-Zeinstra, S. M. (2010). Diagnostic accuracy of history taking and physical examination for assessing anterior cruciate

ligament lesions of the knee in primary care. *Arch Phys Med Rehabil, 91*, 1452–1459.

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., & Group, Q.-S. (2013). A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol, 66*, 1093–1104.

Zhang, W., Doherty, M., Peat, G., Bierma-Zeinstra, M. A., Arden, N. K., Bresnihan, B., Herrero-Beaumont, G., Kirschner, S., Leeb, B. F., Lohmander, L. S., Mazieres, B., Pavelka, K., Punzi, L., So, A. K., Tuncer, T., Watt, I., & Bijlsma, J. W. (2010). EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. *Ann Rheum Dis, 69*, 483–489.