Clinical assessment of effusion in knee osteoarthritis—A systematic review

Nasimah Maricar[a],*, Michael J. Callaghan[a], Matthew J. Parkes[a], David T. Felson[a,b,c], Terence W. O'Neill[a,b]

[a] Arthritis Research UK Centre for Epidemiology, Institute of Inflammation and Repair, Faculty of Medical and Human Sciences, Manchester Academic Health Science Centre, University of Manchester, Manchester M13 9PT, UK
[b] NIHR Manchester Musculoskeletal Biomedical Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK
[c] Clinical Epidemiology Unit, Boston University School of Medicine, Boston, MA

ARTICLE INFO

ABSTRACT

Objective: The aim of this systematic review was to determine the validity and inter- and intra-observer reliability of the assessment of knee joint effusion in osteoarthritis (OA) of the knee.
Methods: MEDLINE, Web of Knowledge, CINAHL, EMBASE, and AMED were searched from their inception to February 2015. Articles were included according to a priori defined criteria: samples containing participants with knee OA; prospective evaluation of clinical tests and assessments of knee effusion that included reliability, sensitivity, and specificity of these tests.
Results: A total of 10 publications were reviewed. Eight of these considered reliability and four on validity of clinical assessments against ultrasound effusion. It was not possible to undertake a meta-analysis of reliability or validity because of differences in study designs and the clinical tests. Intra-observer kappa agreement for visible swelling ranged from 0.37 (suprapatellar) to 1.0 (prepatellar); for bulge sign 0.47 and balloon sign 0.37. Inter-observer kappa agreement for visible swelling ranged from −0.02 (prepatellar) to 0.65 (infrapatellar), the balloon sign −0.11 to 0.82, patellar tap −0.02 to 0.75 and bulge sign kappa −0.04 to 0.14 or reliability coefficient 0.97. Reliability and diagnostic accuracy tended to be better in experienced observers. Very few data looked at performance of individual clinical tests with sensitivity ranging 18.2–85.7% and specificity 35.3–93.3%, both higher with larger effusions.
Conclusion: The majority of unstandardized clinical tests to assess joint effusion in knee OA had relatively low intra- and inter-observer reliability. There is some evidence experience improved reliability and diagnostic accuracy of tests. Currently there is insufficient evidence to recommend any particular test in clinical practice.
© 2016 The Authors. Published by Elsevier HS Journals, Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Knee effusion is common among people with knee osteoarthritis (OA). In those with knee pain and having radiographic OA, nine out of 10 people will have imaging evidence of effusion with 55% having a moderate to large effusion [1]. Clinical assessment for knee effusion is thus an integral part of routine physical examination in knee OA. A range of clinical tests have been used to assess the presence of knee effusion including visible inspection of swelling [2], palpation of the knee [3,4], and a range of dynamic tests including ballottement, patellar tap and the sweep test, though the terminology used in the literature to describe the tests is inconsistent [2,4–7]. To have clinical value and utility, a clinical test should be both valid, in that it should detect an effusion if present and exclude it if not present, and also reliable, meaning that repeated assessments with either the same or different observers result in the same conclusion. Knowledge of the reliability and validity of the currently used clinical tests for knee effusion in knee OA would help inform the optimum approach taken in clinical assessment of these patients and help in informing diagnostic or intervention strategies. To the best of our knowledge there has been no systematic review that pool reliability and validity data from individual studies. We therefore undertook a systematic review with the aim of determining (1) intra- and inter-observer reliability and (2) performance characteristics, of clinical tests for the presence of effusion in knee OA.

* Corresponding author.
  E-mail address: nasimah.maricar@postgrad.manchester.ac.uk (N. Maricar).

Methods

Search strategies

Publications which reported either intra- and inter-observer reliability of clinical assessments of knee effusion or the validity of clinical assessments using imaging such as ultrasound (US) and magnetic resonance imaging (MRI) were identified from searching six databases up to February 2015: Medline (1948 onwards), Embase (1974 onwards), AMED (1985 onwards), Cumulative Index to Nursing and Allied Health Literature (CINAHL) Plus (1937 onwards), Web of Knowledge (1950 onwards), and the Cochrane Central Registers for Controlled Trials. The databases were searched individually for all possible terms and combination of terms to accommodate differences in their search engines. All medical subject-heading searches (MeSH) were exploded when possible. The key terms used in combination ("OR") were test, examination, assessment, physical, clinical, MRI, US, effusion and swelling which were then used in combination ("AND") with knee. Full details of the MEDLINE search strategy appear in Appendix 1. Hand searches were also performed in addition to additional searches through Google Scholar and Reference Manager Search engines. Reference lists of publications were also searched for other relevant publications. There was no language restriction. Eligibility assessment of trials for inclusion in the review was performed unblinded by 1 reviewer (NM).

Study selection criteria

Publications considered were those that included adults with knee OA based on the American College of Rheumatology (ACR) clinical classification criteria for OA [8] or from detailed clinical and/or radiographic assessment of the knee joint. Due to a limited number of studies on knee OA, for this review, studies with mixed samples of OA and inflammatory arthritis (IA) or knee pain were also included. Studies were selected if they reported evaluation of the validity of clinical assessment of knee effusion against imaging assessment of effusion, or the reliability of clinical assessment of effusion undertaken either by the same observer (intra-observer reliability) or different observers (inter-observer reliability).

Quality assessment

For publications evaluating reliability of the clinical tests for knee joint effusion, study quality was graded using the "reliability" section of COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) with close reference to a separate publication which provided detailed description of the scoring system [9]. To evaluate the quality of diagnostic accuracy studies on knee joint effusion, the diagnostic study appraisal checklist from Centre of Evidence-Based Medicine (CEBM) was used. Two assessors (NM, MJC) independently assessed and scored the publications for quality and reached consensus in cases of disagreement.

Data extraction and analyses

The data from the selected papers were extracted independently by two reviewers (NM, MJP) using a standardised form. Information about the type of study, methods evaluated, subject and observer characteristics, assessment protocols and their results including relevant data for calculation of sensitivity, specificity and reliability of the effusion tests were extracted. In this review, "sensitivity" refers to the ability of a test to identify correctly the presence of effusion in individuals with knee effusion and "specificity" the ability of the test to correctly exclude individuals without knee effusion, assessed by an US or MRI as "gold standard." Intra- and inter-observer reliability is a measure of the degree of agreement when the test is repeated by the same observer or between different observers, respectively. The kappa (κ) statistic was used to denote agreement for dichotomous variables while weighted kappa (κ^{ω}) and reliability coefficient (R_c) was used for ordinal and continuous variables, respectively. We did not undertake a formal meta-analysis for reliability and performance characteristics data because of the different methodologies and differences in reporting among the studies. For kappa and R_c, we considered values of less than 0.2 to indicate poor agreement, between 0.21 and 0.40 fair, 0.41–0.60 moderate, 0.61–0.80 as good and values more than 0.80 to indicate excellent agreement [10].

Results

Search outcome

The results of the search are summarized in Figure 1. In total 10 publications met the inclusion criteria; some of these addressed intra- and/or inter-observer reliability only, others reliability and validity (Tables 1 and 2). Several authors were contacted for further information. Where information was obtained it was included in the review [11,12]. Two articles reported results from the same trials and were considered as one study to avoid potential bias from over-reporting/counting [11,13].

Characteristics of included studies

Eight publications evaluated intra- and/or inter-observer reliability testing for clinical tests of effusion [12,14–20]. The number of assessors varied from two to six [12,14–20]. Two studies reported on reliability of skilled rheumatologists, orthopedic specialists and medical consultants [14,18], four on clinicians of mixed experience such as trainees and consultants [15,16,19,20] and one study reported on reliability of physiotherapists with 7–10 years of post-qualification experience [12].

Four publications reported sensitivity and/or specificity of knee effusion tests where the clinical tests were compared against US as the standard [11,18,20,21]. We found one study [22] that compared clinical assessment of effusion against MRI assessment. However, we were unable to include this study as it did not provide sufficient data to allow the determination of sensitivity and specificity and also the clinical test used to assess the knee effusion was not stated. Of the 10 publications included in the review, seven comprised patients with knee OA [11,14–16,19–21], one comprised patients with knee pain (which included those with OA and inflammatory arthritis) [18], one recent presentation to primary care or to rheumatology clinics or on the waiting list for total knee replacement that included patients with a range of knee pain severity and OA [12], and one comprised a population sample including people with knee OA [17].

Clinical tests used in studies

Clinical tests used to assess knee effusion were categorized as either visual inspection, or dynamic testing involving (1) movement of fluid across the knee with the presence of effusion denoted by reappearance of fluid distension (bulge sign) and (2) pressure over the patella with the presence of effusion determined by palpable ballottement or tapping of the patella against the femoral condyle (patellar tap), or palpable distension of the underlying joint line by fluid fluctuation (balloon sign). We looked