Measurement properties of patient-reported outcome measures (PROMS) in Patellofemoral Pain Syndrome

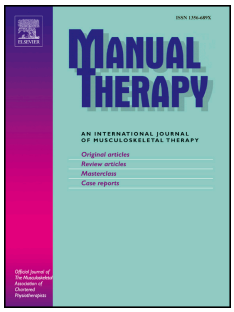
Green, Andrew; Liles, Clive; Rushton, Alison; Kyte, Derek

DOI:
10.1016/j.math.2014.05.013

Document Version
Peer reviewed version

Citation for published version (Harvard):
Green, A, Liles, C, Rushton, A & Kyte, DG 2014, 'Measurement properties of patient-reported outcome measures (PROMS) in Patellofemoral Pain Syndrome: a systematic review', Manual Therapy, vol. 19, no. 6, pp. 517-526. https://doi.org/10.1016/j.math.2014.05.013

Accepted Manuscript



Measurement Properties of Patient-Reported Outcome Measures (PROMS) in Patellofemoral Pain Syndrome: A Systematic Review

Andrew Green, MSc MCSP MMACP Clive Liles, MA MCSP Alison Rushton, EdD, MSc, FMACP Derek G. Kyte, MSc MCSP MMACP

PII: S1356-689X(14)00115-5

DOI: 10.1016/j.math.2014.05.013

Reference: YMATH 1578

To appear in: *Manual Therapy*

Received Date: 23 January 2014

Revised Date: 14 April 2014

Accepted Date: 21 May 2014

Please cite this article as: Green A, Liles C, Rushton A, Kyte DG, Measurement Properties of Patient-Reported Outcome Measures (PROMS) in Patellofemoral Pain Syndrome: A Systematic Review, *Manual Therapy* (2014), doi: 10.1016/j.math.2014.05.013.

Title

Measurement Properties of Patient-Reported Outcome Measures (PROMS) in Patellofemoral Pain Syndrome: A Systematic Review

Keywords

Patellofemoral Pain Syndrome, Patient Reported Outcome Measures, Systematic Review, Measurement Properties.

Authorship

Andrew Green[a,*], MSc MCSP MMACP, Lower Limb Clinical Specialist Physiotherapist.

Clive Liles[b], MA MCSP, Lecturer in Physiotherapy

Dr Alison Rushton[b], EdD, MSc, FMACP, Senior Lecturer in Physiotherapy

Derek G. Kyte[b], MSc MCSP MMACP, Physiotherapy Teaching Associate

Affiliations

[a]The Dudley Group NHS Foundation Trust, Russells Hall Hospital, Dudley, West Midlands, DY1 2HQ.

[b]School of Sport, Exercise and Rehabilitation Sciences, College of Life and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

Corresponding author

Andrew Green

Email: Andrew.Green@dgh.nhs.uk

ABSTRACT

This systematic review investigated the measurement properties of disease-specific patient-reported outcome measures used in Patellofemoral Pain Syndrome. Two independent reviewers conducted a systematic search of key databases (MEDLINE, EMBASE, AMED, CINHAL+ and the Cochrane Library from inception to August 2013) to identify relevant studies. A third reviewer mediated in the event of disagreement. Methodological quality was evaluated using the validated COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) tool. Data synthesis across studies determined the level of evidence for each patient-reported outcome measure. The search strategy returned 2177 citations. Following the eligibility review phase, seven studies, evaluating twelve different patient-reported outcome measures, met inclusion criteria. A 'moderate' level of evidence supported the structural validity of several measures: the Flandry Questionnaire, Anterior Knee Pain Scale, Modified Functional Index Questionnaire, Eng and Pierrynowski Questionnaire and Visual Analogue Scales for 'usual' and 'worst' pain. In addition, there was a 'Limited' level of evidence supporting the test-retest reliability and validity (cross-cultural, hypothesis testing) of the Persian version of the Anterior Knee Pain Scale. Other measurement properties were evaluated with poor methodological quality, and many properties were not evaluated in any of the included papers. Current disease-specific outcome measures for Patellofemoral Pain Syndrome require further investigation. Future studies should evaluate all important measurement properties, utilising an appropriate framework such as COSMIN to guide study design, to facilitate optimal methodological quality.

INTRODUCTION

Patellofemoral pain syndrome (PFPS) is a common knee disorder, with a typical pattern of symptoms characterised by anterior peripatella or retropatella knee pain (Heintjes et al., 2009; Collins et al., 2010; Hossain et al., 2011). Aggravating factors include activities or movements which either increase patellofemoral joint compression and/or produce mechanical forces in the surrounding soft tissue structures; for example: ascending/descending stairs, sitting with a flexed knee for prolonged periods, squatting, running, jumping or kneeling (Witvrouw et al., 2000; Crossley et al., 2002; Barton et al., 2008; Thijs et al., 2008; Tan et al., 2010). As many of these activities are an important part of daily life, PFPS may have a considerable impact on an individual's wellbeing (Collins et al., 2008; Tan et al., 2010). This impact may be especially debilitating as PFPS symptoms often reoccur, becoming chronic (Nimon et al., 1998; Stathopulu and Baildam, 2003; Collins et al., 2008; Boling et al., 2010).

Whilst the aetiology of PFPS is debated, there is some consensus that its development may be secondary to a functional or structural mal-alignment at the patellofemoral joint, or of the lower extremity as a whole (Powers, 2003; Barton et al., 2008; Heintjes et al., 2009; Carry et al., 2010; Hossain et al., 2011). There may be multiple interacting factors which cause mal-alignment, such as muscle strength or timing issues, altered tissue extensibility or bony morphology (Powers, 2003; Barton et al., 2008; Heintjes et al., 2009; Bennell et al., 2010).

Physiotherapy is the most common intervention in PFPS (Crossley et al., 2001; Heintjes et al., 2003), however, there is no clear consensus regarding the optimal components of a management programme. As a consequence, a wide variety of treatment techniques are employed by therapists including: quadriceps

strengthening, vastus medialus obliques (VMO) muscle retraining, biofeedback, hip muscle strengthening, proximal strengthening, spinal manipulation, mobilisation, taping, knee supports, foot orthoses and stretching of the hamstrings, illiotial band, patella retinaculum or anterior hip (Crossley et al., 2002; Iverson et al., 2008; Heintjes et al., 2009; Earl and Hoch, 2011; Hossain et al., 2011; Callaghan and Selfe, 2012). In the absence of guidelines outlining the most favourable PFPS treatment options, physiotherapists should appraise their own management, utilising high quality, disease-specific, PFPS outcome measures to guide and evaluate patient care, so they may deliver efficacious treatment tailored to the individual (DoH, 2010; CSP, 2012; HCPC, 2013).

A number of patient-reported outcome measures (PROMs) have been developed to assess symptoms and function in patients with PFPS. These disease-specific measures are designed to be more sensitive to change in their target population than region-specific measures, which evaluate general knee disorders. When making the choice of which PROM to use in practice, it is important to examine their respective measurement properties, so that the optimal instrument can be confidently employed. These properties should at least satisfy existing minimum standards for PROMs, such as those presented by the International Society for Quality of Life research (Reeve et al., 2013). Previous systematic reviews that have evaluated the measurement properties of knee PROMs, have tended to focus on region-specific measures used in general knee conditions (Bellamy et al., 1997; Sun et al., 1997; Wang et al., 2010), or non-PFPS-specific musculoskeletal disorders (Smith et al., 2008; Howe et al., 2012), and not all reviews have used a validated tool to determine the quality of the included studies, for example, the COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) tool (Mokkink et al., 2010a) or OMERACT (Outcome Measures in Rheumatology) filter (Boers et al., 1998). The purpose of this study was to evaluate the measurement properties of

disease-specific PROMs for PFPS, using a validated measure of methodological quality.

METHODOLOGY

Design

A systematic review of outcomes was conducted according to a pre-defined protocol informed by the PRISMA guidelines (Liberati et al., 2009), the Cochrane Handbook of Systematic Review Interventions (Higgins and Green, 2011) and the COSMIN group (Mokkink et al., 2010b).

Search strategy

The MEDLINE, EMBASE, AMED, CINHAL+ and Cochrane Library electronic databases were searched from inception to August 2013 (the MEDLINE search strategy is presented in Appendix I). All records were downloaded into Endnote© version 15, and duplicates removed. Two authors (DK, CL) independently screened all citations by title/abstract, before retrieving potentially eligible full text articles for review. Disagreements were resolved through discussion, with a third reviewer on hand to mediate if required. The strength of agreement between investigators was established using Cohen's kappa statistic (Cohen, 1960) and interpreted using set criteria (Landis and Koch, 1977). Remaining articles were subjected to a citation search. Finally, a hand-search of all reference lists was conducted.

Identification of eligible studies

Full text original articles were included if they evaluated at least one PROM measurement property (reliability, validity, responsiveness or interpretability (Mokkink et al., 2010a)) in a cohort of PFPS patients. There are no universally agreed diagnostic criteria for PFPS, therefore, this review used criteria employed by several high-quality randomised controlled trials, each demonstrating treatment efficacy in a PFPS cohort (Collins et al., 2008; van Linschoten et al., 2009; Collins et al., 2010). Thus, studies had to include participants that presented with a main complaint of

patellofemoral pain (defined as anterior peripatellar or retropatellar knee pain) with symptoms that were provoked by at least two of the following: prolonged sitting or kneeling, stair walking, running, squatting, hopping, a positive Clarke's sign or grind test, a positive patellar compression test and recognisable painful symptoms on palpation of the patellar facets (Collins et al., 2008; Syme et al., 2009; van Linschoten et al., 2009). Internationally agreed definitions for each measurement property Mokkink et al.,(2010a) informed the eligibility review. Non-English language papers were excluded.

Data Extraction

Two authors (AG, CL) independently extracted data regarding the following measurement properties: reliability, internal consistency, measurement error), validity (including content, construct, criterion and cross-cultural validity), responsiveness and interpretability (Mokkink et al., 2010a). Disagreements were resolved through discussion with the intervention of a third author (DK) if needed.

Measurement Properties

Reliability examines the degree to which a measurement is free from error, and can be considered in three categories: test-retest reliability (the degree to which results can be replicated over time within a stable environment), this can be further divided into inter-rater reliability (between individuals) and intra-rater reliability (within the same individual); internal consistency (correlation between items that are interrelated); and measurement error (systematic and random error within a patient's outcome score that is not attributed to a true change) (Mokkink et al., 2010b). Validity encompasses: content validity (is the PROM an adequate reflection of the construct to be measured); construct validity (how a PROM performs against pre-defined hypotheses); criterion validity (how a PROM compares to a 'gold standard' if available); and cross-cultural validity (how well the translated PROM reflects the

original version) (Mokkink et al., 2010b). Responsiveness is the ability of an outcome measure to detect a clinically meaningful change in a patient's condition over time (Mokkink et al., 2010b). In addition, a PROM must demonstrate adequate interpretability, to ensure that the meaning and significance of changes in score can be easily understood (Mokkink et al., 2010a).

Quality assessment and evidence synthesis

Methodological quality of the included studies was evaluated in order to determine their trustworthiness. Two investigators (AG, CL) independently assessed each study, rating the quality of methods employed to evaluate individual measurement properties, using the validated COSMIN framework (Mokkink et al., 2010a). Disagreements were resolved through discussion with a third author (DK). Papers were rated using a 4-point scale ('poor', 'fair', 'good', 'excellent') (Terwee et al 2012). Synthesis across studies combined findings for each measure and measurement property, taking into account the quality of studies, to determine the level of evidence for each PROM (Schellingerhout et al., 2012). The overall level of evidence was rated as 'strong', 'moderate', 'limited' or 'conflicting', in-line with the criteria proposed by the Cochrane Back Review Group (van Tulder et al., 2003).

RESULTS

Study selection

The search strategy returned 2177 citations. 2155 studies were excluded by title/abstract and 22 full-text articles were retrieved for further review. Of these, 15 full-text articles were excluded as they utilised non-PFPS cohorts, PFPS was not the major complaint of the participants, or because the PFPS diagnostic criteria used by the paper did not meet the defined standards, or was missing altogether. Inter-rater agreement between the investigators during title/abstract screening was 'good' (k=0.68, 95% CI 0.557-0.806) (Cohen, 1998). No additional full-text articles were included following either the citation search or the hand search of reference lists, therefore, 7 papers were included in the final analysis (Figure. 1). The included studies evaluated 12 PROMs, including: the Activity of Daily Living Scale (ADLS) (Irrgang et al., 1998); the Eng and Pierrynowski Questionnaire (EPQ), also known as the Visual Analogue Pain Scale during Activity (Eng and Pierrynowski, 1993); the Flandry Questionnaire (Flandry et al., 1991); the Kujala/Anterior Knee Pain Scale (AKPS) (Kujala et al., 1993); the Modified Functional Index Questionnaire (MFIQ) (Chesworth et al., 1989); the Persian Version Kujala/AKPS (Negahban et al., 2012); the Patellofemoral Function Scale (PFS) (Reid, 1992); the PFPS Severity Scale Syndrome (PSS) (Laprade and Culham, 2002); the Visual Analogue Pain Scale (VAS), also referred to as the Numerical Pain Rating Score (NPRS); and the Visual Analogue Pain Scales for least pain (VAS-L), usual pain (VAS-U) and worst pain (VAS-W).

Study characteristics

Study characteristics are presented in Table 1. The 7 studies examined 384 symptomatic PFPS subjects, largely recruited from the general population, with 1 study recruiting from the military (Laprade and Culham, 2002). The mean age of participants ranged from 23.8 to 32 years old. Average duration of symptoms was reported in 3 papers (Bennell et al., 2000; Crossley et al., 2004; Negahban et al., 2012) and ranged from 12.0 to 38.6 months. No study provided details regarding the severity of participant symptoms. Details of each PROM evaluated across the 7 studies is presented in Table 2. Measurement property data are presented in Table 3. The methodological quality of the studies is presented in Table 4 and the results, presented per PROM, are discussed below.

Activity of Daily Living Scale (ADLS)

Contains 14 items investigating general daily activities and specific functional tasks (e.g. stair descent). Each item is scored 0-5 to provide an overall percentage score. Higher scores indicate better functioning. One study (Piva et al., 2009) found the ADLS responsive, demonstrating a moderate change in score (effect size 0.63). However, this property was evaluated with 'poor' methodological quality.

Eng and Pierrynowski questionnaire (EPQ)

An activity-related pain-rating tool using a 0-10 visual analogue scale. Higher scores indicate more pain. Two studies (Bennell et al., 2000; Crossley et al., 2004) supported reliability ($ICC_{3,1}$ 0.83-0.92), one study (Crossley et al., 2004) found the EPQ responsive (RTE 0.76), and one study (Bennell et al., 2000) reported a Minimal Clinically Important Difference (MCID) of 14 points (23%). However, measurement properties were evaluated with 'poor' methodological quality. Structural validity of the EPQ was supported by Bennell et al. (2000), with a 'moderate' correlation (r=0.66)

with the Flandry questionnaire. This property was evaluated with 'moderate' methodological quality.

Flandry Questionnaire

Consists of 28 visual analogue scale items which investigate the severity of knee symptoms and the ability to perform physical activities. One study (Bennell et al., 2000) supported test-retest reliability ($ICC_{3,1}$ = 0.95) and structural validity (r = 0.66) and also reported a Standard Error of Measurement (SEM) of 34 points (27.6%). However, these measurement properties were evaluated with 'poor' methodological quality.

Kujala/Anterior Knee Pain Scale (AKPS)

A 13-item knee function questionnaire, scored out of 100, with higher scores indicating less disability. Two studies (Bennell et al., 2000; Crossley et al., 2004) supported reliability ($ICC_{3,1}$ = 0.81-0.90) and responsiveness (treatment effect size = 1.15 for responders), however, these measurement properties were evaluated with 'poor' methodological quality. Bennell et al. (2000) reported a moderate correlation for structural validity (r = 0.58); this property was evaluated with 'good' methodological quality.

Modified Functional Index Questionnaire (MFIQ)

Consists of 8 items that measure the ability to perform various functional activities. A maximum score of 16 indicates optimal functioning. Three studies (Chesworth et al., 1989; Bennell et al., 2000; Crossley et al., 2004) supported test-retest reliability (ICC = 0.48; $ICC_{3,1}$ = 0.49-0.94); one study (Harrison et al., 1995) demonstrated a 'very

good' level of internal consistency (pre-treatment α = 0.85 and post-treatment α = 0.88); one study (Bennell et al., 2000) reported the MCID as 2.8 points (16%); two studies (Harrison et al., 1995; Crossley et al., 2004) found the MFIQ responsive with a moderate effect sizes (0.49 and 0.59)., However, all these properties were evaluated with 'poor' methodological quality. Bennell et al. (2000) supported the validity of the MFIQ, evidenced by a moderate correlation (r = -0.66); this property was evaluated with 'good' methodological quality.

The PFPS Severity Scale Syndrome (PSS)

A 10 item visual analogue instrument, examining the effect of PFPS on an individual's functional activities. One study (Laprade and Culham, 2002) supported the test-retest reliability (r = 0.95), however, this property was evaluated with 'poor' methodological quality.

Visual Analogue Scale (VAS)/ Numerical Pain Rating Scale

This scale - scored from 0 (no pain) to 10 (max pain) - evaluates levels of pain. Other versions include: (1) VAS-L (pain at its *least*), (2) VAS-U (*usual* level of pain), and (3) VAS-W (pain when at its *worst)*. Five studies (Chesworth et al., 1989; Harrison et al., 1995; Bennell et al., 2000; Crossley et al., 2004; Piva et al., 2009) supported both test-retest reliability (ICC = 0.56-0.77; $ICC_{3,1}$ = 0.56-0.79) and responsiveness (effect size for improved responders = 0.70-1.22; RTE = 0.95-1.09) of the VAS, VAS-L, VAS-U and VAS-W, however, these properties were evaluated with 'poor' methodological quality in all studies. One study (Bennell et al., 2000), found a minimum change of 3.3cm (33%) on the VAS-U was required to detect a real change in a patient's condition, again methodological quality was 'poor'.

Bennell et al. (2000) established that the VAS-U and VAS-W were moderately correlated (r = 0.63), providing evidence of their structural validity; this property was evaluated with 'good' methodological quality.

The Patellofemoral Function Scale (*PFS)*

Contains 9 items, scored from 0-100, examining both PFPS signs/symptoms and the ability of the patient to perform functional activities; higher scores indicating less disability. One study (Harrison et al., 1995) found the scale was responsive (effect size = 0.81 for responders) and demonstrated a 'minimally acceptable' to 'acceptable' internal consistency (pre-treatment α = 0.65 and post-treatment α = 0.77), however, these properties were evaluated with 'poor' methodological quality.

Persian Version Kujala/AKPS

One study (Negahban et al., 2012) reported 'excellent' test-retest reliability ($ICC_{2,1}$ = 0.96) and confirmed the accuracy of the hypothesis that the Persian Kujala questionnaire would correlate more highly with the SF-36 *physical* questionnaire than the SF-36 *mental* questionnaire (correlation 0.34-0.51 and 0.25-0.37 respectively); these properties were evaluated with 'fair' methodological quality. Negahban et al. (2012) also reported high levels of internal consistency (α = 0.81), however, for this component of the study, methodological quality was 'poor'. Cross-cultural validity was also examined with the authors concluding that no major translation modifications were required, this aspect of the study demonstrated 'good' methodological quality (Negahban et al, (2012).

As this questionnaire was a translated version, it was not synthesised with English language AKPS as the respective findings may not be directly comparable (Schellingerhout et al., 2012), hence presented separately in Table 5.

Synthesis of results across studies

Synthesis of results for each questionnaire with the associated level of evidence is presented in Table 5. There was a 'moderate' level of evidence to support the structural validity of the: EPQ, Flandry Questionnaire, AKPS, MFIQ, VAS-U and VAS-W. In addition, there was a 'limited' level of evidence supporting the reliability (test-retest) and validity (cross-cultural and hypothesis testing) of the Persian version of the AKPS, based on the findings of one paper.

It was not possible to identify supporting evidence for the following PROM measurement properties due to poor methodological quality across the included papers: ADLS (responsiveness); EPQ and AKPS (test-retest reliability, measurement error, responsiveness); Flandry Questionnaire (test-retest reliability, measurement error); MFIS (internal consistency, test-retest reliability, measurement error, responsiveness); PSS (internal consistency); VAS and VAS-L (test-retest reliability, responsiveness); VAS-U (test-retest reliability, measurement error, hypothesis testing, responsiveness); VAS-W (internal consistency, test-retest reliability, responsiveness).

There was no information available for the following PROM measurement properties:

ADLS (internal consistency, test-retest reliability, measurement error, structural validity, hypothesis testing); EPQ and AKPS (internal consistency, hypothesis testing); Flandry Questionnaire (internal consistency, hypothesis testing, responsiveness); MFIQ (hypothesis testing); PSS (test-retest reliability, measurement error, structural validity, hypothesis testing, responsiveness); VAS and VAS-L (internal consistency, measurement error, structural validity, hypothesis testing); VAS-U (internal consistency) and VAS-W (internal consistency, hypothesis testing). In addition, no PROM was examined for the measurement property of interpretability.

DISCUSSION

The objective of this systematic review was to evaluate the measurement properties of disease-specific PROMs for PFPS, to aid clinicians in choosing the best instrument to inform patient management. Unfortunately, the poor methodological quality with which measurement properties were evaluated across the PROMs, makes recommending an optimal instrument problematic.

Principal findings

We found a 'moderate' level of evidence to support the construct validity (structural validity) of six PROMs: the Flandry Questionnaire, AKPS, MFIQ, EPQ, VAS-U and VAS-W. We also found a 'limited' level of evidence supporting the reliability (test-retest) and validity (cross-cultural and hypothesis testing) of the Persian version of the AKPS, based on the findings of one paper. Unfortunately, many other important PROM measurement properties were either evaluated with poor methodological quality (e.g. measurement error), or were not evaluated at all (e.g. interpretability). Common methodological shortcomings included: small sample sizes, absent *a priori* hypotheses, missing details/references for comparator instruments during the

evaluation of responsiveness and a failure to check the uni-dimensionality of a scale prior to the evaluation of internal consistency.

Structural validity, as a component of construct validity, has been identified as a critical element of the overall validity of a PRO instrument (Reeve et al., 2013), it is therefore encouraging that over half of the tools we investigated demonstrated this feature. Unfortunately, no measure was able to satisfy all of the recently agreed minimum standards for PROMs advocated by the International society for Quality of Life research (Reeve et al., 2013).

Comparing these results to those of other authors is difficult. As mentioned previously, there is a lack of systematic reviews focusing on PFPS-specific PROMs used exclusively in PFPS cohorts. Howe et al. (2012) did review the measurement properties of a number of PROMs that arguably could be employed in PFPS, but did so alongside other musculoskeletal disorders, including osteoarthritis, ligament injuries and meniscal lesions. Although the results are not directly comparable, the findings from this study appear similar to ours with regard to the AKPS PROM, which was reviewed in both studies. Using the OMERACT filter, Howe et al. 2012 determined that the tool demonstrated construct validity, however, they also supported its responsiveness, which the current study did not. Our use of the COSMIN tool instead of the OMERACT filter may explain this difference. Finally, Smith and colleagues (2008) evaluated several outcome measures used to assess patellar instability, of which, only the AKPS was included in our study. The findings from both reviews are consistent, namely that poor methodological quality precluded definitive conclusions regarding the measurement properties of the PROMs they investigated.

One of the main purported benefits of disease-specific PROMs is that they may be more sensitive to subtle changes in a patient's condition (i.e. more responsive) than more generic tools (Garratt et al., 2001; Walsh et al., 2003). It is particularly disappointing, therefore, that evidence of responsiveness was lacking in the PROMs we evaluated. Until such time as they are evaluated and validated with greater methodological quality, it is not possible to recommend a disease-specific PROM over an evidence-supported region-specific measure.

Strengths and limitations

A strength of this study is its use of systematic methods to investigate the measurement properties of PROMs employed in PFPS, taking into account the quality of the methods used in the included studies. A limitation is, as there are no universally agreed diagnostic criteria for PFPS, this review used criteria employed by recent high-quality randomised controlled trials. This may have led to the exclusion of some articles that were potentially relevant, but used different diagnostic parameters. Further work is needed to develop definitive PFPS diagnostic criteria.

CONCLUSIONS

Several PROMs used in PFPS demonstrate structural validity including: the Flandry Questionnaire, AKPS, MFIQ, EPQ, VAS-U, and VAS-W. In addition there is limited level of evidence supporting the test-retest reliability and validity (cross-cultural and hypothesis testing) of the Persian version of the AKPS, based on one study. However, no instrument possesses supporting evidence for all important measurement properties (Reeve et al., 2013). The measurement properties of PROMs in PFPS are commonly evaluated with poor methodological quality, and many are yet to be investigated. Current PFPS measures should be subjected to further scrutiny and future studies should evaluate all important measurement properties, utilising an appropriate framework such as COSMIN to guide study design, to facilitate optimal methodological quality.

REFERENCES

Barton CJ, Webster KE, Menz HB. Evaluation of the Scope and Quality of Systematic Reviews on Nonpharmacological Conservative Treatment for Patellofemoral Pain Syndrome. Journal of Orthopaedic & Sports Physical Therapy 2008;38(9):529-41.

Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for a Core Set of Outcome Measures for Future Phase III Clinical Trials in Knee, Hip, and Hand Osteoarthritis. Consensus Development at OMERACT III. The Journal of Rheumatology 1997;24:799-802.

Bennell K, Bartam S, Crossley K, Green S. Outcome measures in patellofemoral pain syndrome: test retest reliability and inter-relationships. Physical Therapy in Sport 2000;1(2):32-41.

Bennell KL, Duncan M, Cowan SM, McConnell J, Hodges P, Crossley KM. Effects of vastus medialis oblique retraining versus general quadriceps strengthening on vasti onset. Medicine & Science in Sports & Exercise 2010;42(5):856-64.

Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT Filter for Outcome Measures in Rheumatology. The Journal of Rheumatology 1998;25(2):198-99.

Boling M, Padua D, Marshall S, Guskiewicz K, Pyne S, Beutler A. Biomechanical risk factors for the development of patellofemoral pain: the JUMP-ACL study... Patellofemoral pain syndrome: proximal, distal, and local factors, an international research retreat, April 30-May 2, 2009, Fells Point, Baltimore, MD. Journal of Orthopaedic & Sports Physical Therapy 2010;40(3):A28-9.

Callaghan M, Selfe J. Patella taping for patellofemroal pain syndrome in adults. Cochrane Database of Systematic Reviews 2012;Issue 4. Art. No.: CD006717. DOI: 10.1002/14651858.CD006717.pub2.

Carry PM, Kanai S, Miller NH, Polousky JD. Adolescent patellofemoral pain: a review of evidence for the role of lower extremity biomechanics and core instability. Orthopedics 2010;33(7):498-507.

Chesworth BM, Culham EG, Tata GE, Peat M. Validation of Outcome Measures in Patients with Patellofemoral Syndrome. Journal of Orthopaedic & Sports Physical Therapy 1989;10(8):302-08.

Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960;20(1):37-46.

Cohen J. Statistical Power Analysis for Social Sciences. 2nd Ed. Hillside, New jersey: Lawrence Erlbaum Associates; 1998

Collins NJ, Crossley KM, Beller E, Darnell R, McPoil T, Vincenzino B. Foot orthoses and physiotherapy in the treatment of patellofemoral pain syndrome: randomised clinical trial. British Medical Journal 2008;337:a1735.

Collins NJ, Crossley KM, Darnell R, Vicenzino B. Predictors of short and long term outcome in patellofemoral pain syndrome: a prospective longitudinal study. BMC Musculoskeletal Disorders 2010;11:11.

Crossley K, Bennell K, Cowan SM, Green S. Analysis of Outcome Measures for Persons With Patellofemoral Pain: Which Are Reliable and Valid? Archives of Physical Therapy and Rehabilitation Medicine 2004;85(5):815-22.

Crossley K, Bennell K, Green S, McConnell J. A Systematic Review of Physical Interventions for Patellofemoral Pain Syndrome. Clinical Journal of Sport Medicine 2001;11(2):103-10.

Crossley KM, Bennell KL, Green S, Cowan SM, McConnell J. Physical therapy for patellofemoral pain. The American Journal of Sports Medicine 2002;30(6):857-65.

CSP. Quality assurance standards for physiotherapy service delivery. Chartered Society of Physiotherapy 2012.
DoH. Equity and excellence: Liberating the NHS. UK Department of Health 2010.
Earl JE, Hoch AZ. A Proximal Strengthening Program Improves Pain, Function, and Biomechanics in Women with Patellofemoral Pain Syndrome. American Journal of Sports Medicine 2011;39(1):154-63.
Eng JJ, Pierrynowski MR. Evaluation of Soft Foot Orthotics in the Treatment of Patellofemoral Pain Syndrome. Physical Therapy 1993;73(2):62-70.
Flandry F, Hunt JP, Terry GC, Hughston J. Analysis of subjective knee complaints using visual analogue scales. American Journal of Sports Medicine 1991;19(2):112-18.
Garratt AM, Klaber Moffett J, Farrin AJ. Responsiveness of Generic and Specific Measures of Health Outcome in Low Back Pain. Spine 2001;26(1):71-77.
Harrison E, Quinney H, Magee D, Sheppard MS, McQuarrie A. Analysis of outcome measures used in the study of patellofemoral pain syndrome. Physiotherapy Canada 1995;47(4):264-72.
HCPC. Standards of proficiency: physiotherapy. Health & Care Professions Council 2013.
Heintjes EM, Berger M, Bierma-Zeinstra SMA, Bernsen RMD, Verhaar JAN, Koes BW. Exercise therapy for patellofemoral pain syndrome (Cochrane review). The Cochrane Library. 4 ed. Chichester, UK: John Wiley & Sons, Ltd; 2003.
Heintjes EM, Berger M, Bierma-Zeinstra SMA, Bernsen RMD, Verhaar JAN, Koes BW. Exercise therapy for patellofemoral pain syndrome. Cochrane Database of Systematic Reviews 2009(4).
Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
Hossain M, Alexander P, Burls A, Jobanputra P. Foot orthoses for patellofemoral pain in adults (Review). Cochrane Database of Systematic Reviews 2011;Issue 1(Art. No.: CD008402. DOI: 10.1002/14651858.CD008402.pub2.).
Howe TE, Dawson LJ, Syme G, Duncan L, Reid J. Evaluation of outcome measures for use in clinical practice for adults with musculoskeletal conditions of the knee: A systematic review. Manual Therapy 2012;17:100-18.
Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. Journal of Bone and Joint Surgery - Series A 1998;80(8):1132-45.
Iverson CA, Sutlive TG, Crowell MS, Morrell RL, Perkins MW, Garber MB, et al. Lumbopelvic manipulation for the treatment of patients with patellofemoral pain syndrome: development of a clinical prediction rule... including commentary by Powers CM with authors' response. Journal of Orthopaedic & Sports Physical Therapy 2008;38(6):297-312.
Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of Patellofemoral Disorders. The Journal of Arthroscopic & Related Surgery 1993;9(2):159-63.
Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. Biometrics 1977;33(1):159-74.
Laprade JA, Culham EG. A self-administered pain severity scale for patellofemoral pain syndrome. Clinical Rehabilitation 2002;16(7):780-8.
Liberati A, Altman DG, Tetzlaff J, Mulrow C, GÃ¸tzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. Annals of Internal Medicine 2009;151(4):W65-W94.
Mokkink L, Terwee C, Gibbons E, Stratford P, Alonso J, Patrick D, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research 2010b;19(4):539-49.

Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. Journal of Clinical Epidemiology 2010a;63(7):737-45.

Negahban H, Pouretezad M, Yazdi M, Sohani SM, Mazaheri M, Salavati M, et al. Persian translation and validation of the kujala Patellofemoral Scale in patients with patellofemoral pain syndrome. Disability and Rehabilitation 2012;34(26):2259-63.

Nimon G, Murray D, Sandow M, Goodfellow J. Natural history of Anterior knee pain: A 14- to 20-year follow-up of nonoperative management. Journal of Pediatric Orthopaedics 1998;18(118-122).

Piva SR, Gil AB, Fitzgerald GK. Responsiveness of the activities of daily living scale of the knee outcome survey and numeric pain rating scale in patients with patellofemoral pain. Journal of Rehabilitation and Medicine 2009;41.

Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint dysfunction: A theoretical perspective. Journal of Orthopaedic & Sports Physical Therapy 2003;33:639-46.

Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. Qual Life Res 2013.

Reid DC. Sports Injury Assessment. New York: Churchill Livingstone; 1992

Schellingerhout J, Verhagen A, Heymans M, Koes B, de Vet H, Terwee C. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. Qual Life Res 2012;21:659-70.

Smith TO, Davies L, O'Driscoll M, Donell ST. An evaluation of the clinical tests and outcome measures used to assess patellar instability. The Knee 2008;15:255-62.

Stathopulu E, Baildam E. Anterior knee pain: a long-term follow-up. Rheumatology 2003;42:380-82.

Sun Y, Sturmer T, Gunther KP, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee - A review of the literature. Clinical rheumatology 1997;16(2):185-98.

Syme G, Rowe P, Martin D, Daly G. Disability in patients with chronic patellofemoral pain syndrome: a randomised controlled trial of VMO selective training versus general quadriceps strengthening. Manual Therapy 2009;14:252-63.

Tan SS, van Linschoten RL, van Middelkoop M, Koes BW, Bierma-Zeinstra SMA, Koopmanschap MA. Cost-utility of exercise therapy in adolescents and young adults suffering from the patelofemoral pain syndrome. Scandinavian Journal of Medicine & Science in Sports 2010;20:568-79.

Thijs Y, De Clercq D, Roosen P, Witvrouw E. Gait-related intrinsic risk factors for patellofemoral pain in novice recreational runners. British Journal of Sports Medicine 2008;42(6):466-71.

van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemsen SP, et al. Supervised exercise therapy versus usual care for patellofemoral pain syndrome: an open label randomised controlled trial. British Medical Journal 2009;339:b4074.

van Tulder M, Furlan AD, Bombardier C, Bouter LM. Updated Method Guidelines for Sytematic Reviews in the Cochrane Collaboration Back Review Group. Spine 2003;28(12):1290-99.

Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a Condition-Specific Instrument for Patients with Low Back Pain/Leg Symptoms Really Necessary?: The Responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. Spine 2003;28(6):607-15.

580 Wang D, Jones MH, Khair MM, Miniaci A. Patient-Reported Outcome Measures for the
581 Knee. The Journal of Knee Surgery 2010;23(3):137-51.
582 Witvrouw E, Lysens R, Bellemans J, Cambier D, Vanderstraeten G. Intrinsic Risk Factors
583 for the Development of Anterior Knee Pain in an Athletic Population. A Two-Year
584 Prospective Study. American Journal of Sports Medicine 2000;28(4):480-89.

585

586

587

588 **Table 1**

589 Characteristics of Included Studies

Study	Country	Subjects (num. of females)	Participants Mean (SD or range)	Duration of Symptoms (months) Mean (SD) unless stated.
Bennell et al., 2000[a]	Australia	50 (33)	23.8 ± 8.9 yrs	17.1 (25.2)
Chesworth et al., 1989[a]	Canada	18 (12)	29.0 yrs (20-50)	not reported
Crossley et al., 2004[a]	Australia	71 (46)	27.5 yrs (14-40)	38.6 (42.6) [Rx Gr] 31.1 (32.2) [Placebo Gr]
Harrison et al., 1995[a]	Canada	56 (7)	24.8 yrs (12-41)	not reported
Laprade and Culham, 2002[b]	Canada	29 (71)	32.0 yrs (20-48)	Range: 3-72
Negahban et al., 2012[a]	Iran	100 (0)	25.3 ± 7.0 yrs	Median & interquartile range: 12 (6-24)
Piva et al., 2009[c]	USA	60 (33)	29.9 ± 9.6 yrs	Distribution: 1-3 (38%); 4-6 (22%), 7-12 (10%), 13-24 (18%), >25 (12%).

590 Settings: [a] = General population, [b] = Military population, [c] = Unknown population.

591 .

592 **Table 2**

593 Characteristics of Patient reported Outcome Measures

Instrument Name	Study ID	Summary of Instrument	Scoring Method
Activity of Daily Living Scale (ADLS)	Piva et al., 2009	14-item scale assessing how patient's knee symptoms affect their ability to perform general daily activities (6 items) and specific functional tasks (8 items).	Each item scored 0-5 (0 = unable; 5 = no difficulty to perform); max score = 70. Percentage calculated (score/70 x 100).
Eng & Pierrynowski Questionnaire/Visual Analogue Pain Scale during Activity (EPQ)	Bennell et al., 2000; Crossley et al., 2004	Visual rating scale that is used to indicate the perception of the level of pain during activity.	10cm horizontal line drawn with annotations along the line for example 'no pain'. Score is the measurement from left hand side to patient's mark.
Flandry Questionnaire	Bennell et al., 2000	Questionnaire to evaluate subjective components of unspecific knee complaints; 28 items relating to severity of symptoms and the ability to perform activities.	Each item (i.e. 28) scored on a VAS (0 – 10); max score = 280.
Kujala/Anterior Knee Pain Scale (AKPS)	Bennell et al., 2000; Crossley et al., 2004	13 item multiple-choice PFPS questionnaire relating to the patient's knee function.	3-5 response choices depending on item; each response allocated a score. Scores vary between 0 and 10; max score = 100.
Modified Functional Index Questionnaire (MFIQ)	Bennell et al., 2000; Chesworth et al., 1989; Crossley et al., 2004; Harrison et al., 1995	8-item questionnaire relating to the ability to perform functional activities on that day.	Each item scored on a 3-point scale: 0 – unable to do; 1 – can do with a problem; 2 – no difficulty; max score = 16.
Persian Version Kujala	Negahban et al., 2012	Translated version of Kujala – Persian (more info see Kujala)	Same as Kujala.
The Patellofemoral Function Scale (PFS)	Harrison et al., 1995.	16 multiple-choice item scale with 9 PROs and 7 CROs. Items are based on pain, ability to perform functional activities and cardinal signs associated with	Items have multiple choice answers (e.g. Jogging – 6 = no restriction; 0 = restricted.) Response scores vary between 0 and 10; max score

		PFPS.	= 100.
The PFPS Severity Scale Syndrome (PSS)	Laprade and Culham, 2002	10 items assessing pain and the ability to perform functional activities.	Each item scored using 10cm VAS; max score = 100.
Visual Analogue Pain Scale (VAS)/ Numerical Pain Rating Score (NPRS)	Chesworth et al., 1989; Piva et al., 2009	Visual rating scale that is used to indicate the perception of the current level of pain.	As in Eng/VAS-A
Visual Analogue Pain Scale when Least pain (VAS-L)	Harrison et al., 1995	Visual rating scale that is used to indicate the perception of the level of pain when at its least.	As in Eng/VAS-A
Visual Analogue Pain Scale when Usual pain (VAS-U)	Bennell et al., 2000; Crossley et al., 2004; Harrison et al., 1995	Visual rating scale that is used to indicate the perception of the usual level of pain i.e. average level of pain.	As in Eng/VAS-A
Visual Analogue Pain Scale when Worst pain (VAS-W)	Bennell et al., 2000; Crossley et al., 2004; Harrison et al., 1995	Visual rating scale that is used to indicate the perception of the level of pain when at its worst.	As in Eng/VAS-A

PFPS – Patellofemoral Pain Syndrome; VAS – Visual Analogue Scale; PRO – Patient-Reported Outcome; CRO – Clinician-Reported Outcome.

Table 3
Results of the measurement properties for the patient reported outcome measures.

Study ID	Study size	Measurement Property: Internal Consistency	Test-Retest Reliability (& Standard Error of Measurement)	Validity	Responsiveness
ADLS					
Piva et al., 2009	n = 60	_	_	_	Effect Size: 1.19(I), 0.03 (NI), 0.63 (overall) Guyatt Index: 1.4 Area under ROC curve: 0.83 MCID = 7.1% (5 pts)
EPQ					
Bennell et al., 2000	n = 50	_	$ICC_{(3,1)}$ = 0.92 Paired *t* test = 2.298 (0.03)[a] SEM: 4.8 (14 points; 23%)	vs Flandry, Pearson r = 0.66 (BC)[S]	_
Crossley et al., 2004	n = 71	_	$ICC_{(3,1)}$ = 0.83 Paired *t* test = -0.10 (0.92)[a]	_	Median score: 1(NI) vs -19(I) RTE = 0.76 RE = No figure provided
Flandry					
Bennell et al., 2000	n = 50	_	$ICC_{(3,1)}$ = 0.95 Paired *t* test = 0.991 (0.33)[a] SEM: 120 (34 points; 27.6%)	vs Eng, Pearson r = 0.66 (BC)[S]	_

Kujala/AKPS					
Bennell et al., 2000	n = 50	_	$ICC_{(3,1)}$ = 0.90 Paired *t* test = -0.673 (0.51)[a] SEM: 4.7 (13 points; 14%)	vs FIQ, Pearson r = 0.58 (BC)[S]	_
Crossley et al., 2004	n = 71	_	$ICC_{(3,1)}$ = 0.81 Paired *t* test = -1.35 (0.20)[a]	_	Median score: 2(NI) vs 15.5(I) RTE = 1.15 RE = 1.24
MFIQ					
Bennell et al., 2000	n = 50	_	$ICC_{(3,1)}$ = 0.94 Paired *t* test = 1.796 (0.09)[a] SEM: 1.0 (2.8 points; 16%)	vs Flandry, Pearson r = -0.66 (BC)[S]	_
Chesworth et al., 1989	n = 18	_	ICC = 0.483	_	ANOVA[a]: F = 21.09; 2,20 df; p<0.001 Newman-Keuls[a]: Pre-Rx: p >0.05 Post-Rx: p<0.01
Crossley et al., 2004	n = 71	_	$ICC_{(3,1)}$ = 0.49 Paired *t* test = -1.34 (0.20)[a]	_	Median score: -0.5(NI) vs 3(I) RTE = 0.49 RE = 0.18
Harrison et al., 1995	n = 56	*Pre-Rx:* Cronbach's α = 0.85 *Post-Rx:* Cronbach's α = 0.88	*Pre-Rx:* Spearman r = 0.69-0.77 *Post-Rx:* Spearman r = 0.84-0.92	_	ANOVA[a]: Pre-Rx: no significant differences (p<0.05) & post Rx: significant differences (p<0.05) between I & NI. Effect Size: *Pre-Rx:* -0.17 Effect Size: *Post-Rx:* 0.59(I), -0.50(NI)
Persian Version Kujala					
Negahban et al., 2012	n =	Cronbach's α =	$ICC_{(2,1)}$ = 0.96	Correlations higher	_

	100	0.81		between Kujala & SF36 physical than Kujala & SF36 mental[H]	
PFS					
Harrison et al., 1995	N = 56	*Pre-Rx:* Cronbach's α = 0.65 *Post-Rx:* Cronbach's α = 0.77			Effect Size: Pre-Rx: no results Effect Size: *Post-Rx:* 0.81(I), -0.31(NI)
PSS					
Laprade and Culham, 2002	n = 29	Spearman r = 0.95	_	_	_
VAS/ NPRS					
Chesworth et al., 1989	n = 18	_	ICC = 0.603	_	ANOVA[a]: F = 19.72; 2,20 df; p< 0.001 Newman-Keuls[a]: Pre-Rx: p >0.05
Piva et al., 2009	n = 60	_	_	_	Effect Size: 1.22(I), 0.26(NI) Guyatt Index: 1.9 Area under ROC curve: 0.84 MCID = 1.2 pts
VAS-L					

Harrison et al., 1995	n = 56	_	*Pre-Rx:* ICC = 0.64 *Post-Rx:* ICC = 0.74	_	ANOVA[a]: Pre-Rx: no significant differences ($p<0.05$) & post Rx: significant differences ($p<0.05$) between I & NI
VAS-U					
Bennell et al., 2000	n = 50	_	$ICC_{(3,1)}$ = 0.77 Paired *t* test = 0.517 (0.61)[a] SEM: 1.2 (3.3cm; 30%)	vs VAS-W, Pearson r = 0.63 (BC)[S]	_
Crossley et al., 2004	n = 71	_	$ICC_{(3,1)}$ = 0.56 Paired *t* test = -0.40 (0.69)[a]	_	Median score: -1(NI) vs -3(I) RTE = 0.95 RE = 1
Harrison et al., 1995	n = 56	_	*Pre-Rx:* ICC = 0.58 *Post-Rx:* ICC = 0.77	_	ANOVA[a]: Pre-Rx: no significant differences ($p<0.05$) & post Rx: significant differences ($p<0.05$) between I & NI Effect Size: *Pre-Rx:* -0.20 Effect Size: *Post-Rx:* 0.75(I), -0.15(NI)
VAS-W					
Bennell et al., 2000	n = 50	_	$ICC_{(3,1)}$ = 0.79 Paired *t* test = 3.301 (0.03)[a] SEM: 1.1 (3.0cm; 30%)	vs VAS-U, Pearson r = 0.63 (BC)[S]	_
Crossley et al., 2004	n = 71	_	$ICC_{(3,1)}$ = 0.76	_	Median score: 0.5 (NI) vs -3.5(I)

Harrison et al., 1995	n = 56	_	Paired *t* test = 1.65 (0.12)[a] *Pre-Rx:* ICC = 0.56 *Post-Rx:* ICC = 0.70	_	RTE = 1.09 RE = No figure provided ANOVA[a]: Pre-Rx: no significant differences (p<0.05) & post Rx: significant differences (p<0.05) between I & NI. Effect Size: *Pre Rx:* 0.02 Effect Size: *Post Rx:* 1.15(I), 0.09(NI)

[S] – Structural validity; [H] – Hypothesis validity.

ROC = Reciever Operating Characteristic; MCID = minimum clinical important difference; ICC = Intraclass Correlation Coefficient; SEM = Standardised Error of Measurement; BC = Best Correlation; NI = Not Improved; I = Improved; RTE = Relative Treatment Effect; RE = Relative Efficiency; ANOVA = A repeated measures analysis of variance; df = degrees of freedom; CI = Confidence Interval; Rx =treatment; WOMAC = Western Ontario and McMaster Universities;

a Statistical Significant Difference (p = 0.05).

Table 4
Methodological quality of each study per measurement property and patient reported outcome measure

	Measurement Property					
	Internal Consistency	**Test-retest Reliability**	**Measurement Error**	**Validity**	**Responsiveness**	**Interpretability**
ADLS						
Piva et al., 2009	_	_	_	_	Poor	_
EPQ						
Bennell et al., 2000	_	Poor	Poor	Good[S]	_	_
Crossley et al., 2004	_	Poor	_	_	Poor	_
Flandry						
Bennell et al., 2000	_	Poor	Poor	Good[S]	_	_
Kujala/AKPS						
Bennell et al., 2000	_	Poor	Poor	Good[S]	_	_
Crossley et al., 2004	_	Poor	_	_	Poor	_
MFIQ						
Bennell et al., 2000	_	Poor	Poor	Good[S]	_	_
Chesworth et al., 1989	_	Poor	_	_	Poor	_

Crossley et al., 2004	_	Poor	_	_	Poor	_
Harrison et al., 1995	Poor	Poor	_	_	Poor	_
Persian Version Kujala*						
Negahban et al., 2012	Poor	Fair	_	Fair[H]; Good[C]	_	_
PFS**						
Harrison et al., 1995	Poor	_	_	_	Poor	_
PSS						
Laprade and Culham, 2002	_	Poor	_	_	_	_
VAS/ NPRS						
Chesworth et al., 1989	_	Poor	_	_	Poor	_
Piva et al., 2009	_	_	_	_	Poor	_
VAS-L						
Harrison et al., 1995	_	Poor	_	_	Poor	_
VAS-U						
Bennell et al., 2000	_	Poor	Poor	Good[S]	_	_
Crossley et al., 2004	_	Poor	_	_	Poor	_

Harrison et al., 1995	_	Poor	_	_	Poor	_
VAS-W						
Bennell et al., 2000	_	Poor	Poor	Good[S]	_	_
Crossley et al., 2004	_	Poor	_	_	Poor	_
Harrison et al., 1995	_	Poor	_	_	Poor	_

[S] – Structural validity; [H] – Hypothesis validity; [C] – Cross-cultural validity
[*] As a translated version cannot be analysed alongside the other PROM
[**] As this measure has components of PRO and CRO measures this cannot be analysed alongside the other PROM

Table 5
Quality of measurement properties per questionnaire

	Measurement Properties					
	Internal Consistency	**Test-retest Reliability**	**Measurement Error**	**Validity**	**Responsiveness**	**Interpretability**
ADLS	nr	nr	nr	nr	?	nr
EPQ	nr	?	?	++[S]	?	nr
Flandry	nr	?	?	++ [S]	nr	nr
Kujala/AKPS	nr	?	?	++ [S]	?	nr
MFIQ	?	?	?	++ [S]	?	nr
PSS	?	nr	nr	nr	nr	nr

VAS/ Numeric Pain Rating Scale	nr	?	nr	nr	?	nr
VAS-L	nr	?	nr	nr	?	nr
VAS-U	nr	?	?	++ [S]	?	nr
VAS-W	nr	?	?	++ [S]	?	nr
Persian Version Kujala [*]	?	+	nr	+ [H,C,‡]	nr	nr
PFS [**]	?	nr	nr	nr	?	nr

Abbreviations: [S] – Structural validity; [H] – Hypothesis validity; [C] – Cross-cultural validity; nr – not reported.
Evidence grading: +++ or --- 'strong' evidence of a positive/negative result, ++ or -- 'moderate' evidence of a positive/negative result, + or - 'limited' evidence of a positive/negative result, ± 'conflicting' evidence, ? unknown due to poor methodological quality.
[‡] Measured against SF 36 physical.
[*] Translated version of Kujala questionnaire.
[**] Measure has PRO and CRO components.

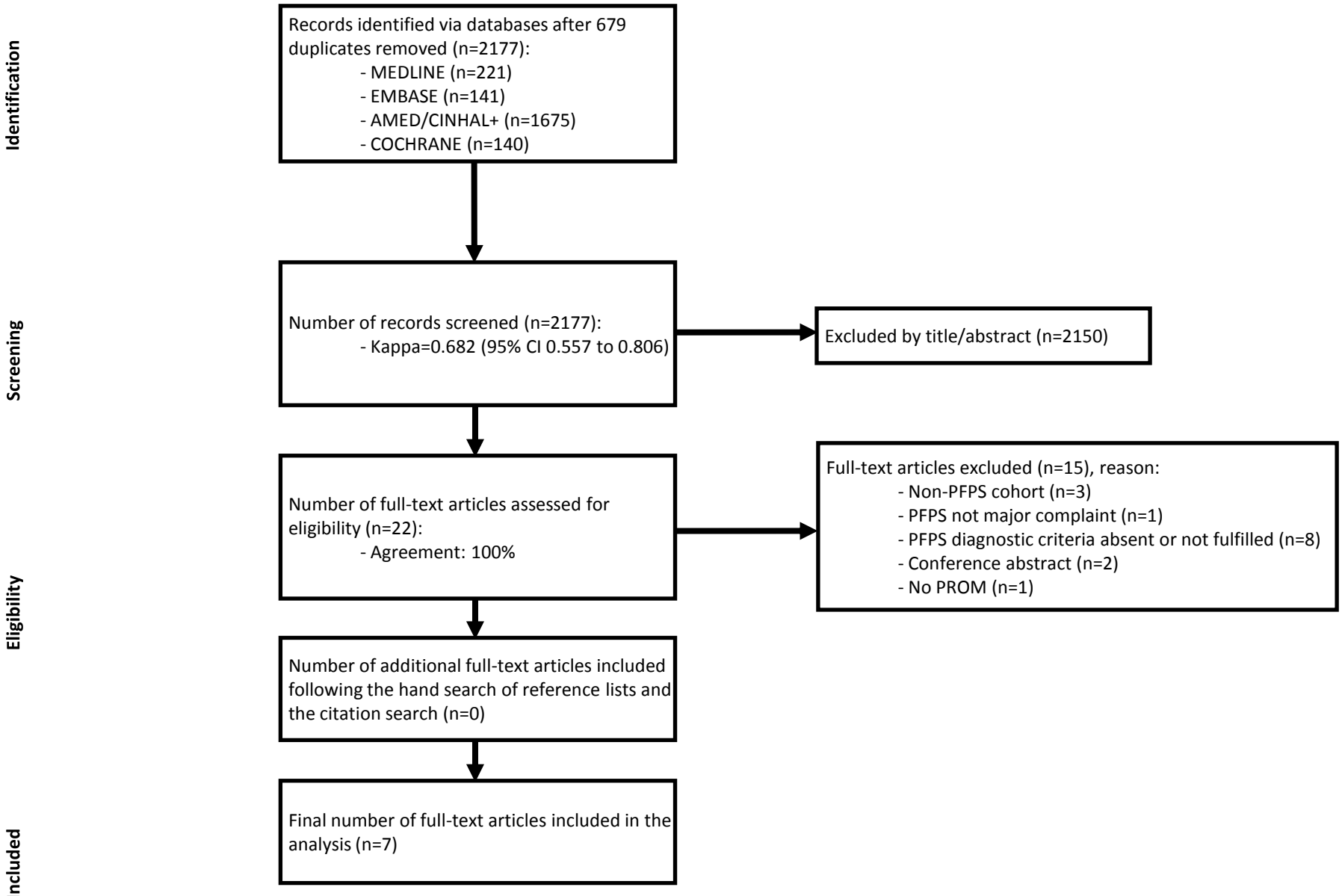


Fig. 1. Flow Diagram of Searches

APPENDIX I

MEDLINE search strategy Aug 2013

1. knee joint or knee or patella or patellofemoral.mp
2. arthralgia or pain.mp
3. Combine 1. And 2.
4. anterior knee pain.mp
5. ((patell$ or femoropatell$ or femoro-patell$ or retropatell$) adj (pain or syndrome or dysfunction)).mp. [mp = title, original title, abstract, name of substance word, subject heading word]
6. ((lateral compression or lateral facet or lateral pressure or odd facet) adj (pain or syndrome or dysfunction)).mp. [mp = title, original title, abstract, name of substance word, subject heading word]
7. ((chondromalac$ or chondropath$) adj (knee$ or patell$ or femoropatell$ or femoro-patell$ or retropatell$)).mp. [mp = title, original title, abstract, name of substance word, subject heading word]
8. or/3-7
9. Clinometric/ or Psychometric.mp (clinomet$/ or psychometr$.mp)
10. (Reliability/ or reliable).mp
11. Validity adj (content or construct or criterion).mp
12. Responsiveness.mp
13. clinical sensitivity.mp
14. Internal adj consistency.mp
1. Measurement adj error.mp
2. Interpretability.mp
17. or/9-16.
18. outcome measur$.mp
19. Questionnair$.mp
20. (patient reported or patient-reported or self-reported) adj (questionnair$ or scale or measure or outcome or outcome measure$).mp
21. (clinician reported or clinician-reported or performance based or performance-based) adj (questionnair$ or scale or measure or outcome ot outcome measur$).mp
22. or/ 18-21
23. 8 and 17 and 22
24. Limit 23 (humans)