

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

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

DOI:

[10.1016/j.math.2014.05.013](https://doi.org/10.1016/j.math.2014.05.013)

License:

Other (please specify with Rights Statement)

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>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

NOTICE: this is the author's version of a work that was accepted for publication. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published 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.

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

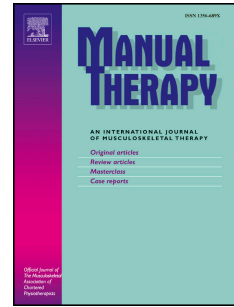
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

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](https://doi.org/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.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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

^aThe Dudley Group NHS Foundation Trust, Russells Hall Hospital, Dudley, West Midlands,
DY1 2HQ.

^bSchool 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.

54
55INTRODUCTION

56 Patellofemoral pain syndrome (PFPS) is a common knee disorder, with a typical
57 pattern of symptoms characterised by anterior peripatella or retropatella knee pain
58 (Heintjes et al., 2009; Collins et al., 2010; Hossain et al., 2011). Aggravating factors
59 include activities or movements which either increase patellofemoral joint
60 compression and/or produce mechanical forces in the surrounding soft tissue
61 structures; for example: ascending/descending stairs, sitting with a flexed knee for
62 prolonged periods, squatting, running, jumping or kneeling (Witvrouw et al., 2000;
63 Crossley et al., 2002; Barton et al., 2008; Thijs et al., 2008; Tan et al., 2010). As
64 many of these activities are an important part of daily life, PFPS may have a
65 considerable impact on an individual's wellbeing (Collins et al., 2008; Tan et al.,
66 2010). This impact may be especially debilitating as PFPS symptoms often reoccur,
67 becoming chronic (Nimon et al., 1998; Stathopulu and Baildam, 2003; Collins et al.,
68 2008; Boling et al., 2010).

69

70 Whilst the aetiology of PFPS is debated, there is some consensus that its
71 development may be secondary to a functional or structural mal-alignment at the
72 patellofemoral joint, or of the lower extremity as a whole (Powers, 2003; Barton et al.,
73 2008; Heintjes et al., 2009; Carry et al., 2010; Hossain et al., 2011). There may be
74 multiple interacting factors which cause mal-alignment, such as muscle strength or
75 timing issues, altered tissue extensibility or bony morphology (Powers, 2003; Barton
76 et al., 2008; Heintjes et al., 2009; Bennell et al., 2010).

77

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

82 strengthening, vastus medialis obliques (VMO) muscle retraining, biofeedback, hip
83 muscle strengthening, proximal strengthening, spinal manipulation, mobilisation,
84 taping, knee supports, foot orthoses and stretching of the hamstrings, illiotial band,
85 patella retinaculum or anterior hip (Crossley et al., 2002; Iverson et al., 2008;
86 Heintjes et al., 2009; Earl and Hoch, 2011; Hossain et al., 2011; Callaghan and Selfe,
87 2012). In the absence of guidelines outlining the most favourable PFPS treatment
88 options, physiotherapists should appraise their own management, utilising high
89 quality, disease-specific, PFPS outcome measures to guide and evaluate patient
90 care, so they may deliver efficacious treatment tailored to the individual (DoH, 2010;
91 CSP, 2012; HCPC, 2013).

92
93 A number of patient-reported outcome measures (PROMs) have been developed to
94 assess symptoms and function in patients with PFPS. These disease-specific
95 measures are designed to be more sensitive to change in their target population than
96 region-specific measures, which evaluate general knee disorders. When making the
97 choice of which PROM to use in practice, it is important to examine their respective
98 measurement properties, so that the optimal instrument can be confidently employed.
99 These properties should at least satisfy existing minimum standards for PROMs,
100 such as those presented by the International Society for Quality of Life research
101 (Reeve et al., 2013). Previous systematic reviews that have evaluated the
102 measurement properties of knee PROMs, have tended to focus on region-specific
103 measures used in general knee conditions (Bellamy et al., 1997; Sun et al., 1997;
104 Wang et al., 2010), or non-PFPS-specific musculoskeletal disorders (Smith et al.,
105 2008; Howe et al., 2012), and not all reviews have used a validated tool to determine
106 the quality of the included studies, for example, the COSMIN (Consensus-based
107 Standards for the Selection of Health Measurement Instruments) tool (Mokkink et al.,
108 2010a) or OMERACT (Outcome Measures in Rheumatology) filter (Boers et al.,
109 1998). The purpose of this study was to evaluate the measurement properties of

110 disease-specific PROMs for PFPS, using a validated measure of methodological
111 quality.
112

ACCEPTED MANUSCRIPT

113 METHODOLOGY

114

115 Design

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

120

121 Search strategy

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

132

133 Identification of eligible studies

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

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

149

150 Data Extraction

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

156

157 Measurement Properties

158 Reliability examines the degree to which a measurement is free from error, and can
159 be considered in three categories: test-retest reliability (the degree to which results
160 can be replicated over time within a stable environment), this can be further divided
161 into inter-rater reliability (between individuals) and intra-rater reliability (within the
162 same individual); internal consistency (correlation between items that are
163 interrelated); and measurement error (systematic and random error within a patient's
164 outcome score that is not attributed to a true change) (Mokkink et al., 2010b).

165 Validity encompasses: content validity (is the PROM an adequate reflection of the
166 construct to be measured); construct validity (how a PROM performs against pre-
167 defined hypotheses); criterion validity (how a PROM compares to a 'gold standard' if
168 available); and cross-cultural validity (how well the translated PROM reflects the

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

174

175 Quality assessment and evidence synthesis

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

187

188

189 RESULTS

190

191 Study selection

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

212

213 Study characteristics

214 Study characteristics are presented in Table 1. The 7 studies examined 384
215 symptomatic PFPS subjects, largely recruited from the general population, with 1
216 study recruiting from the military (Laprade and Culham, 2002). The mean age of
217 participants ranged from 23.8 to 32 years old. Average duration of symptoms was
218 reported in 3 papers (Bennell et al., 2000; Crossley et al., 2004; Negahban et al.,
219 2012) and ranged from 12.0 to 38.6 months. No study provided details regarding the
220 severity of participant symptoms. Details of each PROM evaluated across the 7
221 studies is presented in Table 2. Measurement property data are presented in Table
222 3. The methodological quality of the studies is presented in Table 4 and the results,
223 presented per PROM, are discussed below.

224

225 *Activity of Daily Living Scale (ADLS)*

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

231

232 *Eng and Pierrynowski questionnaire (EPQ)*

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

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

242

243 *Flandry Questionnaire*

244 Consists of 28 visual analogue scale items which investigate the severity of knee
245 symptoms and the ability to perform physical activities. One study (Bennell et al.,
246 2000) supported test-retest reliability ($ICC_{3,1} = 0.95$) and structural validity ($r = 0.66$)
247 and also reported a Standard Error of Measurement (SEM) of 34 points (27.6%).
248 However, these measurement properties were evaluated with 'poor' methodological
249 quality.

250

251 *Kujala/Anterior Knee Pain Scale (AKPS)*

252 A 13-item knee function questionnaire, scored out of 100, with higher scores
253 indicating less disability. Two studies (Bennell et al., 2000; Crossley et al., 2004)
254 supported reliability ($ICC_{3,1} = 0.81-0.90$) and responsiveness (treatment effect size =
255 1.15 for responders), however, these measurement properties were evaluated with
256 'poor' methodological quality. Bennell et al. (2000) reported a moderate correlation
257 for structural validity ($r = 0.58$); this property was evaluated with 'good'
258 methodological quality.

259

260 *Modified Functional Index Questionnaire (MFIQ)*

261 Consists of 8 items that measure the ability to perform various functional activities. A
262 maximum score of 16 indicates optimal functioning. Three studies (Chesworth et al.,
263 1989; Bennell et al., 2000; Crossley et al., 2004) supported test-retest reliability (ICC
264 = 0.48; $ICC_{3,1} = 0.49-0.94$); one study (Harrison et al., 1995) demonstrated a 'very

265 good' level of internal consistency (pre-treatment $\alpha = 0.85$ and post-treatment $\alpha =$
266 0.88); one study (Bennell et al., 2000) reported the MCID as 2.8 points (16%); two
267 studies (Harrison et al., 1995; Crossley et al., 2004) found the MFIQ responsive with
268 a moderate effect sizes (0.49 and 0.59)., However, all these properties were
269 evaluated with 'poor' methodological quality. Bennell et al. (2000) supported the
270 validity of the MFIQ, evidenced by a moderate correlation ($r = -0.66$); this property
271 was evaluated with 'good' methodological quality.

272

273 *The PFPS Severity Scale Syndrome (PSS)*

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

278

279 *Visual Analogue Scale (VAS)/ Numerical Pain Rating Scale*

280 This scale - scored from 0 (no pain) to 10 (max pain) - evaluates levels of pain. Other
281 versions include: (1) VAS-L (pain at its *least*), (2) VAS-U (*usual* level of pain), and (3)
282 VAS-W (pain when at its *worst*). Five studies (Chesworth et al., 1989; Harrison et al.,
283 1995; Bennell et al., 2000; Crossley et al., 2004; Piva et al., 2009) supported both
284 test-retest reliability ($ICC = 0.56-0.77$; $ICC_{3,1} = 0.56-0.79$) and responsiveness (effect
285 size for improved responders = $0.70-1.22$; $RTE = 0.95-1.09$) of the VAS, VAS-L,
286 VAS-U and VAS-W, however, these properties were evaluated with 'poor'
287 methodological quality in all studies. One study (Bennell et al., 2000), found a
288 minimum change of 3.3cm (33%) on the VAS-U was required to detect a real change
289 in a patient's condition, again methodological quality was 'poor'.

290

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

294

295 *The Patellofemoral Function Scale (PFS)*

296 Contains 9 items, scored from 0-100, examining both PFPS signs/symptoms and the
297 ability of the patient to perform functional activities; higher scores indicating less
298 disability. One study (Harrison et al., 1995) found the scale was responsive (effect
299 size = 0.81 for responders) and demonstrated a 'minimally acceptable' to
300 'acceptable' internal consistency (pre-treatment $\alpha = 0.65$ and post-treatment $\alpha =$
301 0.77), however, these properties were evaluated with 'poor' methodological quality.

302

303 *Persian Version Kujala/AKPS*

304 One study (Negahban et al., 2012) reported 'excellent' test-retest reliability ($ICC_{2,1} =$
305 0.96) and confirmed the accuracy of the hypothesis that the Persian Kujala
306 questionnaire would correlate more highly with the SF-36 *physical* questionnaire than
307 the SF-36 *mental* questionnaire (correlation 0.34-0.51 and 0.25-0.37 respectively);
308 these properties were evaluated with 'fair' methodological quality. Negahban et al.
309 (2012) also reported high levels of internal consistency ($\alpha = 0.81$), however, for this
310 component of the study, methodological quality was 'poor'. Cross-cultural validity was
311 also examined with the authors concluding that no major translation modifications
312 were required, this aspect of the study demonstrated 'good' methodological quality
313 (Negahban et al, (2012).

314

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

318

319 Synthesis of results across studies

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

326

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

336

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

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

346 DISCUSSION

347

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

353

354 Principal findings

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

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

366

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

373

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

390

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

398

399 Strengths and limitations

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

407

408

CONCLUSIONS

409

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

421

422

423

424

REFERENCES

- 425 Barton CJ, Webster KE, Menz HB. Evaluation of the Scope and Quality of Systematic
426 Reviews on Nonpharmacological Conservative Treatment for Patellofemoral
427 Pain Syndrome. *Journal of Orthopaedic & Sports Physical Therapy*
428 2008;38(9):529-41.
- 429 Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, et al. Recommendations for
430 a Core Set of Outcome Measures for Future Phase III Clinical Trials in Knee, Hip,
431 and Hand Osteoarthritis. Consensus Development at OMERACT III. *The Journal*
432 *of Rheumatology* 1997;24:799-802.
- 433 Bennell K, Bartam S, Crossley K, Green S. Outcome measures in patellofemoral pain
434 syndrome: test retest reliability and inter-relationships. *Physical Therapy in*
435 *Sport* 2000;1(2):32-41.
- 436 Bennell KL, Duncan M, Cowan SM, McConnell J, Hodges P, Crossley KM. Effects of vastus
437 medialis oblique retraining versus general quadriceps strengthening on vasti
438 onset. *Medicine & Science in Sports & Exercise* 2010;42(5):856-64.
- 439 Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT Filter for Outcome Measures in
440 Rheumatology. *The Journal of Rheumatology* 1998;25(2):198-99.
- 441 Boling M, Padua D, Marshall S, Guskiewicz K, Pyne S, Beutler A. Biomechanical risk
442 factors for the development of patellofemoral pain: the JUMP-ACL study...
443 Patellofemoral pain syndrome: proximal, distal, and local factors, an
444 international research retreat, April 30-May 2, 2009, Fells Point, Baltimore, MD.
445 *Journal of Orthopaedic & Sports Physical Therapy* 2010;40(3):A28-9.
- 446 Callaghan M, Selfe J. Patella taping for patellofemoral pain syndrome in adults. *Cochrane*
447 *Database of Systematic Reviews* 2012; Issue 4. Art. No.: CD006717. DOI:
448 10.1002/14651858.CD006717.pub2.
- 449 Carry PM, Kanai S, Miller NH, Polousky JD. Adolescent patellofemoral pain: a review of
450 evidence for the role of lower extremity biomechanics and core instability.
451 *Orthopedics* 2010;33(7):498-507.
- 452 Chesworth BM, Culham EG, Tata GE, Peat M. Validation of Outcome Measures in Patients
453 with Patellofemoral Syndrome. *Journal of Orthopaedic & Sports Physical*
454 *Therapy* 1989;10(8):302-08.
- 455 Cohen J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological*
456 *Measurement* 1960;20(1):37-46.
- 457 Cohen J. *Statistical Power Analysis for Social Sciences*. 2nd Ed. Hillside, New Jersey:
458 Lawrence Erlbaum Associates; 1998
- 459 Collins NJ, Crossley KM, Beller E, Darnell R, McPoil T, Vincenzino B. Foot orthoses and
460 physiotherapy in the treatment of patellofemoral pain syndrome: randomised
461 clinical trial. *British Medical Journal* 2008;337:a1735.
- 462 Collins NJ, Crossley KM, Darnell R, Vincenzino B. Predictors of short and long term
463 outcome in patellofemoral pain syndrome: a prospective longitudinal study.
464 *BMC Musculoskeletal Disorders* 2010;11:11.
- 465 Crossley K, Bennell K, Cowan SM, Green S. Analysis of Outcome Measures for Persons
466 With Patellofemoral Pain: Which Are Reliable and Valid? *Archives of Physical*
467 *Therapy and Rehabilitation Medicine* 2004;85(5):815-22.
- 468 Crossley K, Bennell K, Green S, McConnell J. A Systematic Review of Physical
469 Interventions for Patellofemoral Pain Syndrome. *Clinical Journal of Sport*
470 *Medicine* 2001;11(2):103-10.
- 471 Crossley KM, Bennell KL, Green S, Cowan SM, McConnell J. Physical therapy for
472 patellofemoral pain. *The American Journal of Sports Medicine* 2002;30(6):857-
473 65.

- 474 CSP. Quality assurance standards for physiotherapy service delivery. Chartered Society
475 of Physiotherapy 2012.
- 476 DoH. Equity and excellence: Liberating the NHS. UK Department of Health 2010.
- 477 Earl JE, Hoch AZ. A Proximal Strengthening Program Improves Pain, Function, and
478 Biomechanics in Women with Patellofemoral Pain Syndrome. *American Journal*
479 *of Sports Medicine* 2011;39(1):154-63.
- 480 Eng JJ, Pierrynowski MR. Evaluation of Soft Foot Orthotics in the Treatment of
481 Patellofemoral Pain Syndrome. *Physical Therapy* 1993;73(2):62-70.
- 482 Flandry F, Hunt JP, Terry GC, Hughston J. Analysis of subjective knee complaints using
483 visual analogue scales. *American Journal of Sports Medicine* 1991;19(2):112-18.
- 484 Garratt AM, Klaber Moffett J, Farrin AJ. Responsiveness of Generic and Specific Measures
485 of Health Outcome in Low Back Pain. *Spine* 2001;26(1):71-77.
- 486 Harrison E, Quinney H, Magee D, Sheppard MS, McQuarrie A. Analysis of outcome
487 measures used in the study of patellofemoral pain syndrome. *Physiotherapy*
488 *Canada* 1995;47(4):264-72.
- 489 HCPC. Standards of proficiency: physiotherapy. Health & Care Professions Council 2013.
- 490 Heintjes EM, Berger M, Bierma-Zeinstra SMA, Bernsen RMD, Verhaar JAN, Koes BW.
491 Exercise therapy for patellofemoral pain syndrome (Cochrane review). The
492 Cochrane Library. 4 ed. Chichester, UK: John Wiley & Sons, Ltd; 2003.
- 493 Heintjes EM, Berger M, Bierma-Zeinstra SMA, Bernsen RMD, Verhaar JAN, Koes BW.
494 Exercise therapy for patellofemoral pain syndrome. *Cochrane Database of*
495 *Systematic Reviews* 2009(4).
- 496 Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Interventions* Version
497 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011.
- 498 Hossain M, Alexander P, Burls A, Jobanputra P. Foot orthoses for patellofemoral pain in
499 adults (Review). *Cochrane Database of Systematic Reviews* 2011; Issue 1 (Art.
500 No.: CD008402. DOI: 10.1002/14651858.CD008402.pub2.).
- 501 Howe TE, Dawson LJ, Syme G, Duncan L, Reid J. Evaluation of outcome measures for use
502 in clinical practice for adults with musculoskeletal conditions of the knee: A
503 systematic review. *Manual Therapy* 2012;17:100-18.
- 504 Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-
505 reported measure of function of the knee. *Journal of Bone and Joint Surgery -*
506 *Series A* 1998;80(8):1132-45.
- 507 Iverson CA, Sutlive TG, Crowell MS, Morrell RL, Perkins MW, Garber MB, et al.
508 Lumbopelvic manipulation for the treatment of patients with patellofemoral
509 pain syndrome: development of a clinical prediction rule... including
510 commentary by Powers CM with authors' response. *Journal of Orthopaedic &*
511 *Sports Physical Therapy* 2008;38(6):297-312.
- 512 Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of
513 Patellofemoral Disorders. *The Journal of Arthroscopic & Related Surgery*
514 1993;9(2):159-63.
- 515 Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data.
516 *Biometrics* 1977;33(1):159-74.
- 517 Laprade JA, Culham EG. A self-administered pain severity scale for patellofemoral pain
518 syndrome. *Clinical Rehabilitation* 2002;16(7):780-8.
- 519 Liberati A, Altman DG, Tetzlaff J, Mulrow C, GÅtzsche PC, Ioannidis JPA, et al. The
520 PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of
521 Studies That Evaluate Health Care Interventions: Explanation and Elaboration.
522 *Annals of Internal Medicine* 2009;151(4):W65-W94.
- 523 Mokkink L, Terwee C, Gibbons E, Stratford P, Alonso J, Patrick D, et al. The COSMIN
524 checklist for assessing the methodological quality of studies on measurement
525 properties of health status measurement instruments: an international Delphi
526 study. *Quality of Life Research* 2010b;19(4):539-49.

- 527 Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. The COSMIN study
528 reached international consensus on taxonomy, terminology, and definitions of
529 measurement properties for health-related patient-reported outcomes. *Journal*
530 *of Clinical Epidemiology* 2010a;63(7):737-45.
- 531 Negahban H, Pouretzad M, Yazdi M, Sohani SM, Mazaheri M, Salavati M, et al. Persian
532 translation and validation of the kujala Patellofemoral Scale in patients with
533 patellofemoral pain syndrome. *Disability and Rehabilitation* 2012;34(26):2259-
534 63.
- 535 Nimon G, Murray D, Sandow M, Goodfellow J. Natural history of Anterior knee pain: A
536 14- to 20-year follow-up of nonoperative management. *Journal of Pediatric*
537 *Orthopaedics* 1998;18(118-122).
- 538 Piva SR, Gil AB, Fitzgerald GK. Responsiveness of the activities of daily living scale of the
539 knee outcome survey and numeric pain rating scale in patients with
540 patellofemoral pain. *Journal of Rehabilitation and Medicine* 2009;41.
- 541 Powers CM. The influence of altered lower-extremity kinematics on patellofemoral joint
542 dysfunction: A theoretical perspective. *Journal of Orthopaedic & Sports Physical*
543 *Therapy* 2003;33:639-46.
- 544 Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL
545 recommends minimum standards for patient-reported outcome measures used
546 in patient-centered outcomes and comparative effectiveness research. *Qual Life*
547 *Res* 2013.
- 548 Reid DC. *Sports Injury Assessment*. New York: Churchill Livingstone; 1992
- 549 Schellingerhout J, Verhagen A, Heymans M, Koes B, de Vet H, Terwee C. Measurement
550 properties of disease-specific questionnaires in patients with neck pain: a
551 systematic review. *Qual Life Res* 2012;21:659-70.
- 552 Smith TO, Davies L, O'Driscoll M, Donell ST. An evaluation of the clinical tests and
553 outcome measures used to assess patellar instability. *The Knee* 2008;15:255-62.
- 554 Stathopulu E, Baildam E. Anterior knee pain: a long-term follow-up. *Rheumatology*
555 2003;42:380-82.
- 556 Sun Y, Sturmer T, Gunther KP, Brenner H. Reliability and validity of clinical outcome
557 measurements of osteoarthritis of the hip and knee - A review of the literature.
558 *Clinical rheumatology* 1997;16(2):185-98.
- 559 Syme G, Rowe P, Martin D, Daly G. Disability in patients with chronic patellofemoral pain
560 syndrome: a randomised controlled trial of VMO selective training versus
561 general quadriceps strengthening. *Manual Therapy* 2009;14:252-63.
- 562 Tan SS, van Linschoten RL, van Middelkoop M, Koes BW, Bierma-Zeinstra SMA,
563 Koopmanschap MA. Cost-utility of exercise therapy in adolescents and young
564 adults suffering from the patellofemoral pain syndrome. *Scandinavian Journal of*
565 *Medicine & Science in Sports* 2010;20:568-79.
- 566 Thijs Y, De Clercq D, Roosen P, Witvrouw E. Gait-related intrinsic risk factors for
567 patellofemoral pain in novice recreational runners. *British Journal of Sports*
568 *Medicine* 2008;42(6):466-71.
- 569 van Linschoten R, van Middelkoop M, Berger MY, Heintjes EM, Verhaar JA, Willemssen SP,
570 et al. Supervised exercise therapy versus usual care for patellofemoral pain
571 syndrome: an open label randomised controlled trial. *British Medical Journal*
572 2009;339:b4074.
- 573 van Tulder M, Furlan AD, Bombardier C, Bouter LM. Updated Method Guidelines for
574 Systematic Reviews in the Cochrane Collaboration Back Review Group. *Spine*
575 2003;28(12):1290-99.
- 576 Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a Condition-Specific Instrument for
577 Patients with Low Back Pain/Leg Symptoms Really Necessary?: The
578 Responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine*
579 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

ACCEPTED MANUSCRIPT

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

594 PFPS – Patellofemoral Pain Syndrome; VAS – Visual Analogue Scale; PRO –
595 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 <i>t</i> test = 2.298 (0.03) ^a SEM: 4.8 (14 points; 23%)	vs Flandry, Pearson <i>r</i> = 0.66 (BC) ^S	–
Crossley et al., 2004	n = 71	–	ICC _(3,1) = 0.83 Paired <i>t</i> 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 <i>t</i> test = 0.991 (0.33) ^a SEM: 120 (34 points; 27.6%)	vs Eng, Pearson <i>r</i> = 0.66 (BC) ^S	–

Kujala/AKPS

Bennell et al., 2000	n = 50	–	ICC _(3,1) = 0.90 Paired <i>t</i> test = -0.673 (0.51) ^a SEM: 4.7 (13 points; 14%)	vs FIQ, Pearson <i>r</i> = 0.58 (BC) ^S	–
Crossley et al., 2004	n = 71	–	ICC _(3,1) = 0.81 Paired <i>t</i> 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 <i>t</i> test = 1.796 (0.09) ^a SEM: 1.0 (2.8 points; 16%)	vs Flandry, Pearson <i>r</i> = -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 <i>t</i> 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	<i>Pre-Rx</i> : Cronbach's α = 0.85 <i>Post-Rx</i> : Cronbach's α = 0.88	<i>Pre-Rx</i> : Spearman <i>r</i> = 0.69- 0.77 <i>Post-Rx</i> : Spearman <i>r</i> = 0.84- 0.92	–	ANOVA ^a : <i>Pre-Rx</i> : no significant differences (p<0.05) & <i>post Rx</i> : significant differences (p<0.05) between I & NI. Effect Size: <i>Pre-Rx</i> : -0.17 Effect Size: <i>Post-Rx</i> : 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 =
0.77Effect Size: *Pre-Rx*: no results
Effect Size: *Post-Rx*: 0.81(I), -0.31(NI)**PSS**Laprade and Culham,
2002n = 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 < 0.001Newman-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 <i>t</i> test = 0.517 (0.61) ^a SEM: 1.2 (3.3cm; 30%)	vs VAS-W, Pearson <i>r</i> = 0.63 (BC) ^S	—
Crossley et al., 2004	n = 71	ICC _(3,1) = 0.56 Paired <i>t</i> 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: <i>Pre-Rx</i> : -0.20 Effect Size: <i>Post-Rx</i> : 0.75(I), -0.15(NI)
VAS-W				
Bennell et al., 2000	n = 50	ICC _(3,1) = 0.79 Paired <i>t</i> test = 3.301 (0.03) ^a SEM: 1.1 (3.0cm; 30%)	vs VAS-U, Pearson <i>r</i> = 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 <i>t</i> test = 1.65 (0.12) ^a	RTE = 1.09
		<i>Pre-Rx</i> : ICC = 0.56	RE = No figure provided
		<i>Post-Rx</i> : ICC = 0.70	ANOVA ^a : <i>Pre-Rx</i> : no significant differences ($p < 0.05$) & <i>post Rx</i> : significant differences ($p < 0.05$) between I & NI.
			Effect Size: <i>Pre Rx</i> : 0.02
			Effect Size: <i>Post Rx</i> : 1.15(I), 0.09(NI)

^S – Structural validity; ^H – Hypothesis validity.

ROC = Receiver 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.

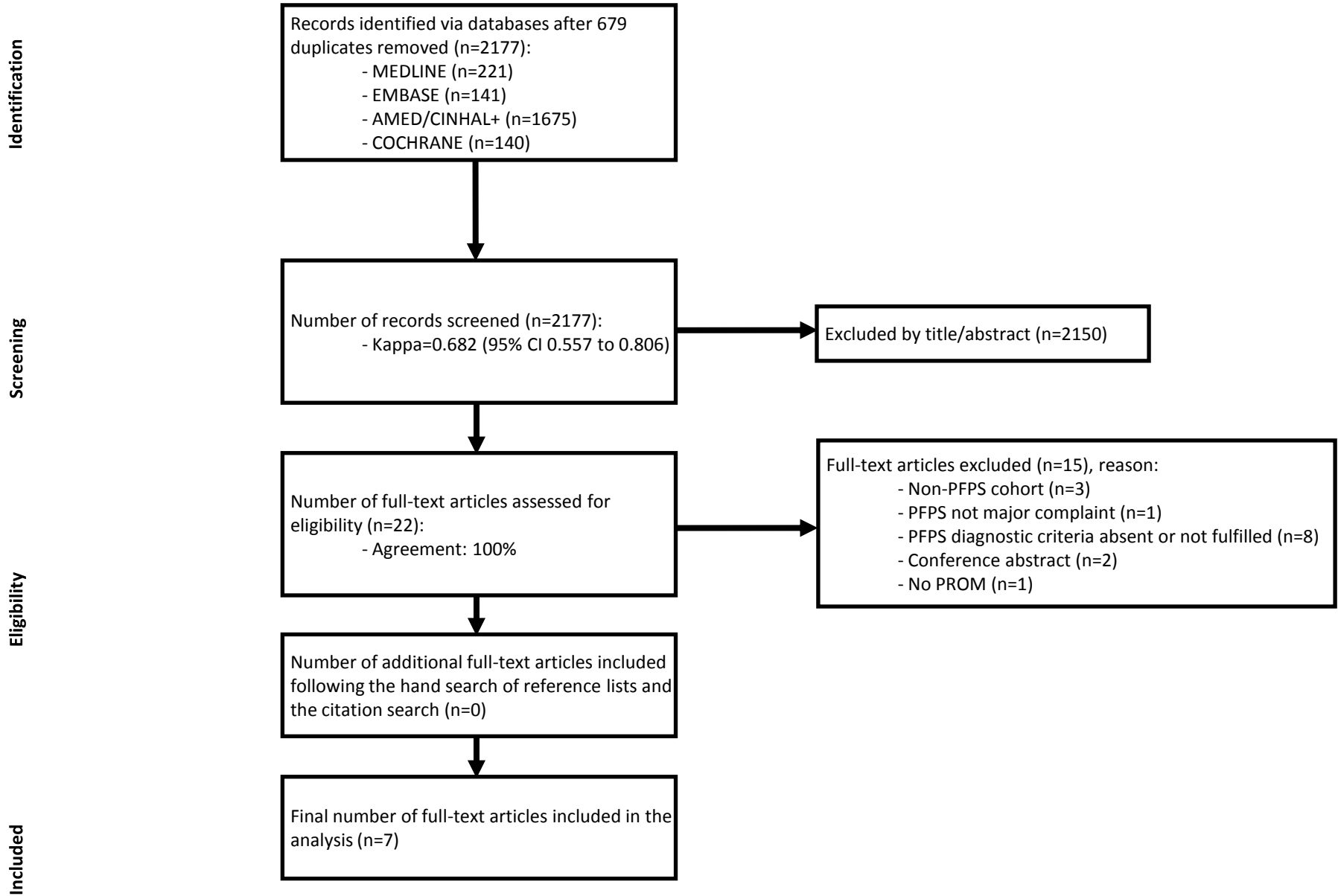


Fig. 1. Flow Diagram of Searches

APPENDIX IMEDLINE 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 or outcome measur\$).mp
22. or/ 18-21
23. 8 and 17 and 22
24. Limit 23 (humans)