Knee Surgery, Sports Traumatology, Arthroscopy
https://doi.org/10.1007/s00167-021-06625-5

KNEE

Patient-reported outcome scores following patellar instability surgery-high prevalence does not equal high responsiveness: a systematic review

Justin A. Magnuson[1] · Brooks N. Platt[1] · Anthony J. Zacharias[1] · Lucy C. Bowers[1] · Cale A. Jacobs[1] · Joseph N. Liu[2] · Austin V. Stone[1]

Received: 14 January 2021 / Accepted: 25 May 2021
© European Society of Sports Traumatology, Knee Surgery, Arthroscopy (ESSKA) 2021

Abstract

Purpose The purpose of this study was to determine the utilization and responsiveness of common patient-reported outcomes (PROs) in patients undergoing surgery for patellar instability.

Methods Using PRISMA guidelines, a systematic review of studies reporting outcomes following surgical intervention for patellar instability was conducted using Pubmed, Cochrane, OVID Medline, and Google Scholar. Subgroup analysis of articles reporting at least two PROs with baseline and follow-up data were used to evaluate responsiveness of instruments using relative efficiency and effect size.

Results From the search, 2,848 unique articles were found, of which 178 were included in final analysis (7,122 patients, mean age 22.6, 63.6% female). The most commonly used PRO was the Kujala score (79.2%), followed by the Lysholm (34.8%), and Tegner (30.9%). Seventy-nine articles were eligible for subgroup analysis. The Kujala had a higher relative efficiency than ten of the 14 instruments to which it was compared but had lower relative efficiency compared to the IKDC and Lysholm scores. The Banff Patella Instability Instrument (BPII) and the Norwich score, condition-specific tools, were unable to be fully assessed due to rarity of use and lack of comparisons.

Conclusion The hypothesis that the Kujala score is the most commonly used PRO for patellar instability, although other instruments offer greater efficiency was supported by our results. The IKDC and Lysholm scores had similar effect sizes but higher relative efficiencies than the Kujala, thus suggesting better responsiveness. This analysis adds useful information for surgeons on the effectiveness of the most common PRO's for evaluating patellofemoral instability outcomes.

Level of evidence Level III.

Keywords Knee · Patient-reported outcomes · Patellar instability · Medial patellofemoral ligament reconstruction

Introduction

Patellar dislocation is a common knee injury with incidences ranging from 5.8 per 100,000 in the general population to 33 per 100,000 in adolescent females [10]. Research on patellar instability surgical outcomes is an emerging domain, yet clinical outcomes data remains relatively inconsistent [24]. There is no consensus on objective clinical data and patient-reported outcome Measures (PROs) reporting after patellar instability procedures. While first time patellar dislocations are traditionally treated nonoperatively, recent debate supports potential early surgical intervention if the recurrent instability of the patella score risk criteria are met. [12, 27] Because young, active patients are at significantly higher risk for repeat dislocation [10], effective outcomes reporting are necessary to determine the benefit of surgical treatment.

Variability in reporting methodology in the literature and lack of objective responsiveness data among available PRO instruments leads to difficulty reaching consensus guidelines. Responsiveness refers to an instruments ability to detect change over time given an intervention [2, 8, 25].

✉ Austin V. Stone
austinvstonemd@gmail.com

[1] Division of Sports Medicine, Department of Orthopaedic Surgery and Sports Medicine, University of Kentucky, 740 S. Limestone, K401, Lexington, KY 40503, USA

[2] USC Epstein Family Center for Sports Medicine, Keck Medicine of USC, Los Angeles, CA, USA

Published online: 16 June 2021

To be useful, it is critical that an instrument be responsive in the range of outcomes in which clinically significant changes are likely to occur [17]. A more responsive PRO would thus be able to detect more differences in patient outcome from a pre-intervention to post-intervention state. The purpose of this systematic review was to evaluate the variability in outcomes reported after patellar instability procedures and compare the responsiveness of commonly used PRO instruments. It was hypothesized that Kujala score, a common general knee outcome score [7] would be the most common instrument used, however, it would be less responsive than condition specific tools. Such information on the effectiveness of the most common PRO's used for patellofemoral outcomes would be useful for surgeons in determining the optimal method to evaluate post-treatment pain, function, and satisfaction.

Materials and methods

A comprehensive literature review was performed in searchable database listings (PubMed, Cochrane, OVID Medline, and Google Scholar) using the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines to identify all articles pertaining to patellar instability. Articles were identified using the keywords: medial patellofemoral ligament, medial patellofemoral ligament reconstruction, MPFL, medial patellofemoral complex, tibial tubercle osteotomy, Fulkerson, Elmslie Trillat, Maquet. Cross-referencing was performed to identify any potentially missed articles. Inclusion criteria were English-language publication, any study with subjective or objective clinical outcomes after patellar instability surgery and standard deviation or standard error measurements for the respective PROs. Exclusion criteria included basic science or cadaveric studies, review articles, meta-analyses, case reports, and editorials, studies focusing primarily on imaging without clinical correlations, and studies focusing on diagnostic or screening techniques.

Several metrics were collected for all studies meeting the inclusion criteria. Basic study descriptors included year and journal of publication, level of evidence, number of patients, number of knees studied, patient demographics, and mean follow-up. Outcomes recorded included revision rate and other complications. Articles reporting at least two PROs at a minimum of 1-year follow-up were included in responsiveness calculations. For this, pre- and post-operative means and standard deviations of PROs were recorded. Requiring at least two PROs per manuscript allowed for relative efficiency comparisons to be made.

Quality assessment

Each article was assigned a level of evidence based on the Oxford Centre for Evidence-Based Medicine [11]. The Methodological Index for Nonrandomized Studies (MINORS) tool was used to evaluate the quality of the included non-randomized studies studies [28]. The MINORS tool uses 12 factors rated on a scale of 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate) to evaluate study quality. The randomized studies included were evaluated using the Modified Coleman Methodology score [6].

Included studies were independently assessed by four authors (JM, AZ, LB, and BP), and any disagreements were resolved by arbitration and consensus.

Statistical analysis

The responsiveness was compared between PRO instruments by calculating effect sizes, standardized response means, and the relative efficiency of each outcome tool as previously described [20, 21, 25]. Effect size is a measure of the magnitude of the pre- to post-operative change accounting for the variability of a given PRO instrument. Effect sizes are defined as small if they fall between 0.2 and 0.49, moderate if between 0.5 and 0.79, and large if above 0.8 [5, 9]. We then calculated the relative efficiency (RE) to directly compare the responsiveness between two different PROs instruments that were used within the same patient population. When comparing two PRO tools, a relative efficiency value less than one indicates that the first PRO tool is less responsive than the other tool [21, 25]. In the current analysis, relative efficiency values less than 0.80 were defined as being indicative of poorer responsiveness, values between 0.80 and 1.20 were indicative of equal responsiveness between tools, and values greater than 1.20 were defined as being indicative of greater responsiveness. To avoid misinterpretation due to small sample sizes, we did not include PRO tools in the responsiveness analyses if the tool appeared in only one article.

Results

We identified 177 studies that met our inclusion/exclusion criteria with a total of 7,176 patients (6,513 knees at final analysis, mean age = 22.7 years (Fig. 1). The mean number of knees per article was 37 (range, 5–298 knees), and the mean follow-up was 48.9 months (range, 6–300 months) (Table 1). The most common procedure used was MPFL reconstruction (130 studies, 5,881 procedures). Other

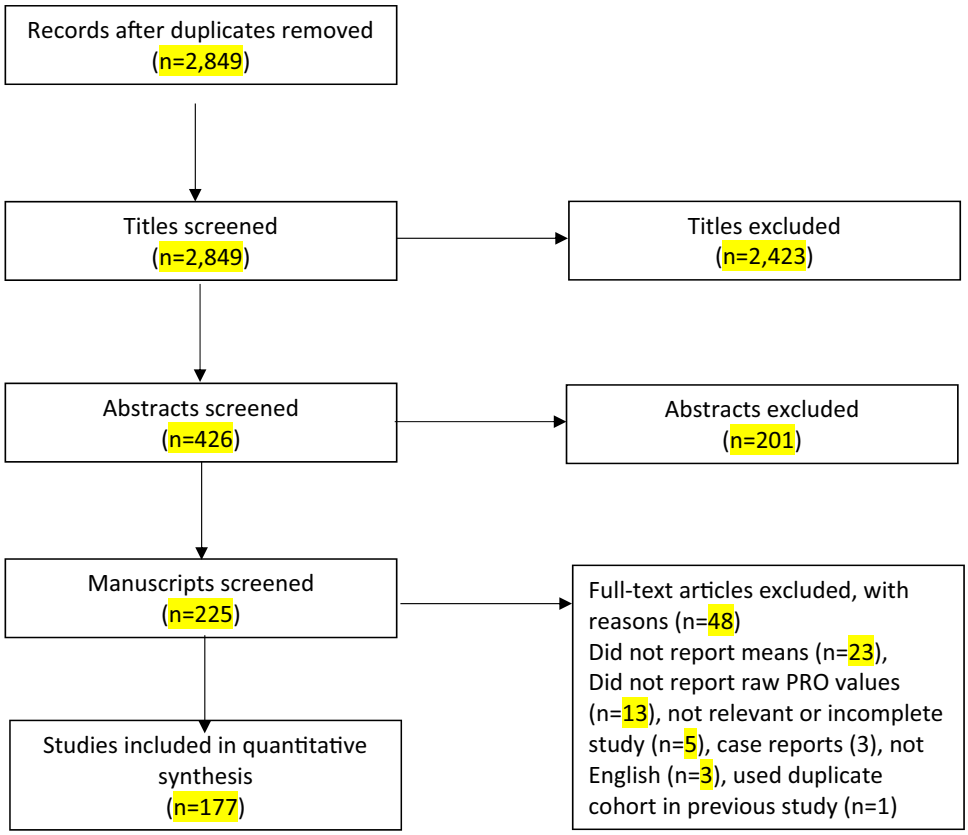

Fig. 1 PRISMA diagram

Table 1 Summative data of the included studies

Study characteristics	Results (range)	Studies reporting (%)
Mean no. of patients	41 (5–298)	177 (100)
Mean no. of knees	37 (5–298)	177 (100)
Mean percent of males	32.4 (0–90.9)	171 (96.6)
Mean age, yr	22.7 (8–74)	170 (96.0)
Mean follow-up, mo	48.9 (6–300)	158 (89.2)
Mean MINORS score of non-randomized studies	11.6 (6–24)	169 (–)
Mean Coleman score of randomized studies	53.3	8 (–)
Reported radiographic measures	–	126 (71.2)
Reported revision surgery	–	129 (72.8)

common procedures included tibial tubercle osteotomy (19 studies, 509 procedures), lateral release (16 studies, 332 procedures), and Elmslie-Trillat (12 studies, 350 procedures). Non-operative treatment was reported in 4 studies for 118 knees (Table 2). The mean level of evidence was 3.3, with 48 of 177 (27.1%) classified as level 1 or 2 evidence. The mean minors score was 12/24 and the mean Coleman score for the randomized studies was 53/65 (Table 1).

Radiographic measures were reported in 126 of 177 studies (71.2%) (Table 1). The TT-TG distance (78/126), patellar tilt (54/126), sulcus angle (48/126), and congruence angle (48/126) were the most common measures, while fewer studies reported the Caton-Deschamp index (41/126) and

Table 2 Common procedure breakdown

Procedure	Number of studies	Number of procedures
MPFL reconstruction	130	5881
Tibial tubercle osteotomy	19	509
Lateral release	16	332
Elmslie-Trillat	12	350
Trochleoplasty	7	172
MPFL repair	7	125
Fulkerson	6	148
Non-operative	4	118

Insall-Salvetti index (38/126). Reoperation was reported in 129 articles (72.9%). PROs were reported in all articles and 54 different PRO instruments were identified. The mean number of PROs per article was 2.76 (range, 1–14), and 125 of 177 (70.6%) used 2 or more PROs. The most commonly used PRO was the Kujala ($n=142$ studies), followed by the Lysholm ($n=62$), Tegner ($n=55$), International Knee Documentation Committee (IKDC, $n=38$), and visual analogue scale (VAS, $n=24$). Other commonly used PRO instruments included the Knee Injury and Osteoarthritis Outcome Score-Quality of Life (KOOS-QOL, $n=15$), KOOS-Pain Score (KOOS-PS, $n=14$), KOOS-Symptoms (KOOS-Sx, $n=14$), KOOS-Activities of Daily Living (KOOS-ADL, $n=13$), KOOS-Sport ($n=13$), Fulkerson ($n=9$), and Short Form-12 Physical Component Score (SF-12 PCS, $n=4$).

From the full set of 177 articles, we identified a subset of 80 articles that met the inclusion criteria for responsiveness analyses. The highest effect size was attained by the Banff Patellar Instability Instrument (BPII) (Table 3). The Kujala had a relative efficiency greater than 10 of the 14 instruments to which it was compared but had lower RE compared to the IKDC and Lysholm scores (Table 4).

Discussion

The principal finding of this study is data collection and outcomes reporting are highly variable in the patellar instability literature. More than 50 different PRO instruments were used. The most common instrument used was the Kujala, while the most responsive instrument analyzed was the Cincinnati knee score, and the score with the highest RE was the Lysholm. Our findings of high variability in the patellar instability literature consisting of multiple PRO instruments of different responsiveness highlights the challenges in comparing outcomes between studies.

This finding is not unique to patellar instability studies, as variable outcomes reporting was previously identified in rotator cuff, shoulder instability, anterior cruciate ligament, and hip arthroscopy literature [16, 22, 23, 32]. More than 20 PROs were reported after hip arthroscopy procedures [32], 16 different PROs were used after rotator cuff procedures [16], and 16 were found in ACL outcomes literature [23]. Consistent reporting is especially important for patellar instability because patients are usually young, active, and have the potential to lose significant function if not managed properly [10].

Pooled data from multiple studies are an effective strategy to identify subsets of patients who may be at risk for inferior outcomes; however, the variability in outcomes measured and the specificity of the common scores used limits our ability to do so. For example, many studies did not report pre-operative PRO scores, standard deviations, or routine radiographic findings. A number of studies also did not report important information such as follow-up time or mean age. Revision rates were also reported only approximately 75% of the time, which limits the assessment of the survivorship of any procedures in question.

The majority of articles in the current review (80%) used the Kujala score, which evaluates symptoms related to

Table 3 PROs included in subgroup analysis and effect size

PRO	Number of studies	Total number of patients	Preoperative mean	Postoperative mean	Effect size
Kujala	74	3076	56.9	85.8	1.7
Lysholm	38	1609	53.5	88.0	2.4
IKDC	18	1089	50.4	81.5	1.9
Tegner	20	900	4.0	4.9	0.5
VAS	12	451	4.8	3.4	0.6
KOOS-QOL	5	285	34.9	64.7	1.1
KOOS-PS	5	285	55.8	63.9	0.3
KOOS-SX	5	229	62.3	78.1	0.7
KOOS-ADL	4	213	70.0	85.9	0.6
KOOS-Sport	4	213	35.2	61.9	0.8
BPII	4	724	26.6	67.1	2.9
Cincinnati	3	111	51.1	90.2	2.3
Fulkerson	3	107	52.1	86.5	1.7
SF-12 Physical	2	66	44.3	60.8	1.4
NRS rest	2	212	2.2	1.6	0.3
NRS activity	2	212	3.2	3.6	0.2

IKDC International Knee Documentation Committee Subjective Knee Form; *VAS* Visual Analogue Scale; *KOOS* Knee Injury and Osteoarthritis Outcome Score; *ADL* Activities of Daily Life; *SF-12* 12-Item Short Form Survey Physical Component Summary; *NRS* Numeric Rating System

Table 4 Relative efficiency of individual comparisons of PROs (values > 1 indicate that PRO in left column has greater relative efficiency compared to PRO in top row; values < 1 indicate that PRO in left column has lesser relative efficiency than PRO in top row; "—" indicates that PROs were unable to be compared as they were not used together in any studies analyzed)

	Kuj	T	L	I	V	KQ	KP	KSx	KA	KSp	F	SFP	Nr	Na	C
Kuj		0.3	1.5	1.3	0.4	1.0	0.1	0.6	0.5	0.8	0.7	1.3	0.1	0.0	0.7
T	2.9		8.6	11.9	---	---	---	---	---	---	---	---	---	---	---
L	0.7	0.1		0.8	0.2	---	---	---	---	---	1	1.1	---	---	---
I	0.8	0.1	1.2		0.3	---	---	---	---	---	---	---	---	---	---
V	2.8	---	4.8	3.0		---	---	1.0	---	---	---	---	---	---	---
KQ	1.0	---	---	---	---		0.1	0.6	0.5	0.8	---	---	---	---	---
KP	16.1	---	---	---	---	17.6		0.9	0.8	0.9	---	---	---	---	---
KSx	1.7	---	---	---	1.0	1.7	1.1		0.9	1.3	---	---	---	---	---
KA	1.9	---	---	---	---	1.9	1.2	1.1		1.5	---	---	---	---	---
KSp	1.3	---	---	---	---	1.3	1.2	0.8	0.7		---	---	---	---	---
F	1.5	---	1.0	---	---	---	---	---	---	---		---	---	---	---
SFP	0.8	---	1.0	---	---	---	---	---	---	---	---		---	---	---
Nr	13.5	---	---	---	---	---	---	---	---	---	---	---		0.5	---
Na	28.9	---	---	---	---	---	---	---	---	---	---	---	2.1		---
C	1.4	---	---	---	---	---	---	---	---	---	---	---	---	---	

Kuj Kujala; *T* Tegner; *L* Lysholm; *I* International Knee Documentation Committee; *V* Visual Analogue Scale; *KQ* KOOS QOL; *KP* KOOS pain; *KA* KOOS ADL; *KSp* KOOS Sport; *F* Fulkerson; *SFP* Short Form–12 Physical Component Score; *Nr* numerical rating scale rest; *Na* numerical rating scale activity; *C* Cincinnati

patellofemoral disorders but focuses on anterior knee pain [18]. Though the instrument was initially developed, as a valuable tool for assessing outcomes related to anterior knee pain, only 1 of its 13 questions directly addresses symptoms of patellar instability [18]. The adaptation of the PRO instrument to evaluate patellar instability, which is associated with pain but is not often the principle complaint, limits its effectiveness when compared to instruments specifically developed to measure patellar instability outcomes.

When assessing patients with patellar instability, multiple common knee PROs, such as the Kujala, IKDC, Lysholm, and Fulkerson, have demonstrated significant ceiling effects [4, 7, 15, 26]. Ceiling effects are especially important to consider in the young population who have recurrent patellar instability because they have the potential for large improvements. In younger more active patients and athletes, baseline functional scores may be high which predisposes them to ceiling effects and an inability to achieve the minimal clinically important difference. For example, a high functioning athlete who sustained a patellar dislocation and was unable to participate at an elite level prior to surgery, may have high functional scores on less sensitive outcomes measures and appear to derive little benefit following surgery despite a successful post-operative return to sport. This lack of sensitivity is a ceiling effect and demonstrates why proper outcomes measurements and return to activity questionnaires offer valuable insights into surgical effectiveness.

In addition to ceiling effects, the IKDC, Lysholm, and KOOS scores have been shown to not significantly increase after 6 months following ACL reconstruction [1]. A timeline of maximal subjective outcome improvement shorter than the time expected to return to sport for most athletes demonstrates the limitation of these scales in detecting clinically significant improvements in athletes. Though the Tegner scales has not been associated with ceiling effects [3], its responsiveness was found to be particularly low in our analysis. This finding is likely due to the one-dimensional nature of the tool, as it only consists of one 10-point scale based on activity and large majority return to a high level of activity after MPFL reconstruction [33].

While comparisons were limited, the general health-related instrument the SF-12 PCS was about as responsive as the knee-specific instruments studies. Such a finding demonstrates that while this general instrument does not fully assess the specific condition of the patient, it may be useful as a general outcome measure for this population. In contrast, pain scales, the VAS and NRS, were much less responsive than the SF-12 and knee-specific instruments.

Overall, the results of this review provide guidance as to which of the general knee and general health PRO instruments are the most responsive in the patellar dislocation population. The results also display the lack of use of instruments specifically designed to assess patellar dislocation outcomes: the BPII and the Norwich scales. The BPII is a quality of life score introduced in 2013 and modified in 2016 specifically to evaluate patellar instability [13, 14, 19]. Unlike other scales, such as the Kujala, IKDC, and Lysholm, ceiling effects have not been demonstrated in this instrument [14, 19]. The Norwich Patellar Instability Score is a symptom score developed in 2014 [31]. While its use in the literature is limited thus far, its responsiveness is to comparable to the Lysholm score, and it has the advantage of being specifically validated in first-time dislocators [29] and without ceiling effects [30].

The reporting of outcomes after patellar instability was found to be highly variable, with more than 50 different PRO instruments used. As such, comparison across studies is currently difficult. Use of a standardized set of measures to assess outcomes after patellar dislocation would offer the potential to draw stronger conclusions from pooled data. On the basis of our comparative responsiveness results and previously reported properties of the different PRO instruments in a critical review by Hiemstra et al. [15] we recommend more common use of specific instruments such as the BPII and Norwich scores. While their rarity of use does not allow for a recommendation of widespread adoption at this juncture, their specificity of design [13, 14, 19, 31] offers an opportunity to assess the outcomes of patellar instability patients more accurately. Future studies using these scores in tandem with other more common knee outcome scores would allow for an assessment of their ES and RE, thus moving toward a more universal system of measuring patellar instability outcomes. To judge the efficacy of surgical techniques addressing patellar instability, consistent and reliable outcome reporting is necessary across studies.

Our study has the common limitations of systematic reviews. All procedures used to treat patellar instability were reviewed for the purposes of this study. The breadth of procedures examined was not narrowed because the goal was not to judge treatment efficacy. Instead, we sought to identify which outcome measures were being reported and to assess their responsiveness. For the same reason, we did not specify patient demographics or age-ranges. Many articles were also excluded from the responsiveness analyses as only 73 of 151 articles reported pre- and post-operative means and SDs of at least 2 PROs required for those calculations. While this limited our ability to perform a responsiveness analysis for all PROs, the criteria aimed to preserve the integrity of our study and limit small sample size effects. More widespread reporting of means and standard deviations would allow for better comparisons to be made between studies and tools.

Conclusion

The results showed wide variability in the patellofemoral literature with regards to PRO use. The hypothesis that the Kujala score in the most commonly used PRO for patellar instability was supported by the data, although other instruments offer greater efficiency. The BPII and the Norwich are patellar instability specific instruments but were only used in two studies and one study, respectively.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00167-021-06625-5.

Funding Not applicable.

Availability of data and material Available via Pubmed and Cochrane Databases.

Declarations

Conflict of interest Dr. Stone: AANA—committee member AOSSM—committee member; Research Support Flexion therapeutics, Allosource; Unpaid consultant—Smith & Nephew, Allosource; Grant from Arthrex; Education payments from Medwest Associates; Hospitality payments from Wright Medical Technology; Travel and Lodging Smith and Nephew. Dr. Liu, Education and Travel and Lodging from Smith & Nephew Inc., Education for Arthrex Travel and Lodging and Food and beverage for Exatech Inc. Dr. Jacobs, Consultant for Flexion Therapeutics HH, Research funding from Flexion Therapeutics and Smith & Nephew.

Ethical approval Not Applicable.

References

1. Agarwalla A, Puzzitiello RN, Liu JN, Cvetanovich GL, Gowd AK, Verma NN et al (2019) Timeline for maximal subjective outcome improvement after anterior cruciate ligament reconstruction. Am J Sports Med 47:2501–2509
2. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al (1997) 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. J Rheumatol 24:799–802
3. Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR (2009) The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. Am J Sports Med 37:890–897
4. Cerciello S, Corona K, Morris BJ, Visona E, Maccauro G, Maffulli N et al (2018) Cross-cultural adaptation and validation of the Italian versions of the Kujala, Larsen, Lysholm and Fulkerson scores in patients with patellofemoral disorders. J Orthop Traumatol 19:18
5. Cohen J (1977) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Eribaum, New Jersey, pp 8–17
6. Coleman BD, Khan KM, Maffulli N, Cook JL, Wark JD (2000) Studies of surgical outcome after patellar tendinopathy: clinical significance of methodological deficiencies and guidelines for future studies. Victorian Institute of Sport Tendon Study Group. Scand J Med Sci Sports 10:2–11
7. Dammerer D, Liebensteiner MC, Kujala UM, Emmanuel K, Kopf S, Dirisamer F et al (2018) Validation of the German version of the Kujala score in patients with patellofemoral instability: a prospective multi-centre study. Arch Orthop Trauma Surg 138:527–535
8. Davies GM, Watson DJ, Bellamy N (1999) Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care Res 12:172–179
9. Dunlop WCJ, Vaslow J, Burke M (1996) Meta-analysis of experiments with matched groups or repeated measures designs. Psychol Methods 1:170–177
10. Fithian DC, Paxton EW, Stone ML, Silva P, Davis DK, Elias DA et al (2004) Epidemiology and natural history of acute patellar dislocation. Am J Sports Med 32:1114–1121
11. Group OLoEW. The Oxford 2011 levels of evidence. https://www.cebm.net/index.aspx?o=5653
12. Hevesi M, Heidenreich MJ, Camp CL, Hewett TE, Stuart MJ, Dahm DL et al (2019) The recurrent instability of the patella score: a statistically based model for prediction of long-term recurrence risk after first-time dislocation. Arthroscopy 35:537–543
13. Hiemstra LA, Kerslake S, Lafave M, Mohtadi NG (2016) Concurrent validation of the Banff Patella instability instrument to the Norwich Patellar Instability Score and the Kujala Score in patients with patellofemoral instability. Orthop J Sports Med 4:2325967116646085
14. Hiemstra LA, Kerslake S, Lafave MR, Heard SM, Buchko GM, Mohtadi NG (2013) Initial validity and reliability of the Banff Patella Instability Instrument. Am J Sports Med 41:1629–1635
15. Hiemstra LA, Page JL, Kerslake S (2019) Patient-reported outcome measures for patellofemoral instability: a critical review. Curr Rev Musculoskelet Med 12:124–137
16. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ (2015) A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. Am J Sports Med 43:2572–2582
17. King M, Dobson A (2000) Estimating the responsiveness of an instrument using more than two repeated measures. Biometrics 56:1197–1203
18. Kujala UM, Jaakkola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O (1993) Scoring of patellofemoral disorders. Arthroscopy 9:159–163
19. Lafave MR, Hiemstra L, Kerslake S (2016) Factor analysis and item reduction of the Banff Patella Instability Instrument (BPII): introduction of BPII 2.0. Am J Sports Med 44:2081–2086
20. Lam ET, Lam CL, Fong DY, Huang WW (2013) Is the SF-12 version 2 Health Survey a valid and equivalent substitute for the SF-36 version 2 Health Survey for the Chinese? J Eval Clin Pract 19:200–208
21. Liang MH, Larson MG, Cullen KE, Schwartz JA (1985) Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. Arthritis Rheum 28:542–547
22. Lukenchuk J, Sims LA, Shin JJ (2017) Variability in outcome reporting for operatively managed anterior glenohumeral instability: a systematic review. Arthroscopy 33:477–483
23. Makhni EC, Padaki AS, Petridis PD, Steinhaus ME, Ahmad CS, Cole BJ et al (2015) High variability in outcome reporting patterns in high-impact ACL literature. J Bone Joint Surg Am 97:1529–1542
24. Moatshe G, Cinque ME, Kruckeberg BM, Chahla J, LaPrade RF (2017) Medial patellar instability: a systematic review of

the literature of outcomes after surgical treatment. Arthroscopy 33:1587–1593
25. Nilsdotter AK, Roos EM, Westerlund JP, Roos HP, Lohmander LS (2001) Comparative responsiveness of measures of pain and function after total hip replacement. Arthritis Rheum 45:258–262
26. Paxton EW, Fithian DC, Stone ML, Silva P (2003) The reliability and validity of knee-specific and general health instruments in assessing acute patellar dislocation outcomes. Am J Sports Med 31:487–492
27. Shubin Stein BE, Gruber S, Brady JM (2018) MPFL in first-time dislocators. Curr Rev Musculoskelet Med 11:182–187
28. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J (2003) Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 73:712–716
29. Smith TO, Chester R, Hunt N, Cross JL, Clark A, Donell ST (2016) The Norwich Patellar Instability Score: validity, internal consistency and responsiveness for people conservatively-managed following first-time patellar dislocation. Knee 23:256–260
30. Smith TO, Choudhury A, Navratil R, Hing CB (2019) Psychometric properties of the Norwich Patellar Instability Score in people with recurrent patellar dislocation. Knee 26:1192–1197
31. Smith TO, Donell ST, Clark A, Chester R, Cross J, Kader DF et al (2014) The development, validation and internal consistency of the Norwich Patellar Instability (NPI) score. Knee Surg Sports Traumatol Arthrosc 22:324–335
32. Stone AV, Jacobs CA, Luo TD, Meadows MC, Nho SJ, Stubbs AJ et al (2018) High degree of variability in reporting of clinical and patient-reported outcomes after hip arthroscopy. Am J Sports Med 46:3040–3046
33 Tegner Y, Lysholm J (1985) Rating systems in the evaluation of knee ligament injuries. Clin Orthop Relat Res 198:43–49

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.