

# Common Physical Examination Tests for Patellofemoral Instability Demonstrate Weak Inter-Rater Reliability



Laurie A. Hiemstra, M.D., Ph.D., F.R.C.S.C., Catherine L. O'Brien, M.D., F.R.C.S.C., Mark R. Lafave, CAT(C), Ph.D., and Sarah Kerslake, M.Sc. B.Phty.

**Purpose:** The purpose of this study was to assess the inter-rater reliability of 3 common physical examination assessments, the Q-angle, J-sign, and apprehension test, used to evaluate patients presenting with recurrent lateral patellofemoral instability. **Methods:** A consecutive sample of 38 subjects with recurrent lateral patellofemoral instability in 2013 were included in this reliability study. Two orthopaedic surgeons performed the physical examination maneuvers blinded to each other. The physical examination tests were performed bilaterally and included the Q angle, the J-sign, and the apprehension test. To measure the inter-rater reliability, an intraclass correlation coefficient (ICC 2, $\kappa$ ) was used for both legs independently for Q-angle measurements. A Cohen's kappa was used to measure the inter-rater reliability for the J-sign and the apprehension test. **Results:** The measurement of the Q-angle demonstrated poor reliability (ICC 2, $\kappa$  0.18–0.44). The assessment of the J-sign demonstrated fair to moderate reliability ( $\kappa = 0.31 - 0.41$ ), and the evaluation of apprehension demonstrated fair to substantial reliability ( $\kappa = 0.30 - 0.65$ ). All 3 clinical tests demonstrated substantial variability comparing the reliability on the right and left limbs. Assessment of the quality of apprehension demonstrated good agreement between the examiners. **Conclusions:** In this study of patients with recurrent lateral patellofemoral instability the common physical examination tests, Q-angle, J-sign, and apprehension demonstrated weak inter-rater reliability. These results indicate that these tests are not reliable for communication between health care practitioners or as evaluations for clinical research. **Level of Evidence:** Level IV, therapeutic case series.

Dislocation of the patella is a common injury in both adolescent and adult populations,<sup>1</sup> and subsequent recurrent patellofemoral instability can lead to significant disability and reduced quality of life.<sup>2</sup> The diagnosis of patellofemoral instability can be challenging and, similar to many orthopaedic conditions, is dependent on a detailed history, comprehensive

physical examination, and accurate diagnostic imaging. Physical examination of the unstable patellofemoral joint is not an unequivocal component of the diagnostic process because a myriad of clinical tests and measurements have been described in the literature.<sup>3–6</sup> The most commonly described tests are the Q-angle, the J-sign, and the apprehension test.<sup>7</sup> Unfortunately, clear definitions or reference standards for these tests are lacking, and, despite their widespread use, minimal research has been published regarding validity, reliability, sensitivity, or specificity.<sup>4</sup> In addition, when investigations of reliability and validity have been undertaken, variations in technique and methodological quality may affect the reported diagnostic accuracy. Reproducible and reliable physical examination assessments are necessary, not only for diagnosis but also to inform an educated and evidence-based approach to the management of patellofemoral instability.

The purpose of this study was to assess the inter-rater reliability of 3 common physical examination assessments, the Q-angle, J-sign, and apprehension test, used to evaluate patients presenting with recurrent lateral patellofemoral instability. The study hypothesis was

From Banff Sport Medicine, Banff, Canada (L.A.H., S.K.); Department of Surgery, University of Calgary, Calgary, Canada (L.A.H.); and Department of Health and Physical Education, Mount Royal University, Calgary, Canada (M.R.L.).

The authors report that they have no conflicts of interest in the authorship and publication of this article. Full ICMJE author disclosure forms are available for this article online, as [supplementary material](#).

Received August 19, 2020; accepted January 12, 2021.

Address correspondence to Laurie Anne Hiemstra, M.D., Ph.D., F.R.C.S.C., P.O. Box 1300, Banff, Alberta Canada T1L 1B3. E-mail: [hiemstra@banffsportmed.ca](mailto:hiemstra@banffsportmed.ca)

© 2021 THE AUTHORS. Published by Elsevier Inc. on behalf of the Arthroscopy Association of North America. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>). 2666-061X/201439

<https://doi.org/10.1016/j.asmr.2021.01.004>

that the Q-angle, J-sign, and apprehension tests would not demonstrate strong inter-rater reliability in patients with patellofemoral instability.

## Methods

Consecutive patients presenting with patellofemoral instability in 2013 were identified. The inclusion criterion was a confirmed diagnosis of recurrent patellofemoral instability by one of the assessors. The exclusion criteria were a previous medio-patellofemoral ligament reconstruction, multi-ligament pathology, or worker's compensation cases. A total of 46 patients were assessed, with 5 excluded for previous medio-patellofemoral ligament reconstruction, 2 excluded for multi-ligament knee injury, and 1 excluded as a worker's compensation case. Two orthopaedic surgeons, one an experienced fellowship-trained sport medicine and arthroscopy specialist with an interest in patellofemoral instability (L.A.H.), and the other a fellow undertaking specialty training (C.O.B.), were the raters for this study. The raters performed the history and physical examination maneuvers blinded to each other's findings. The initial rater varied, with 21 patients evaluated first by the experienced surgeon (L.A.H.) and 17 patients first by the fellow (C.O.B.), for a total of 38 patients. The sample size was estimated based on achieving moderate inter-rater reliability of 0.6 to 0.8 and was determined to require 30 subjects.<sup>8</sup>

The physical examination tests were performed bilaterally and included the Q angle, the J-sign, and the apprehension test. The Q-angle was measured with the knee supported on a foam roll at 30° of flexion so that the patella was engaged in the trochlea. Using a long-armed goniometer, the angle formed by lines from the anterior superior iliac spine to the center of the patella and to the tibial tuberosity was measured in degrees. The J-sign was assessed in the sitting position with the hip and knee at 90° of flexion. The patient was asked to extend and flex their knee at moderate speed several times while the tracking of the patella in the trochlear groove was observed by the rater. The J-sign was graded as positive or negative, as well as on a 5-point Likert scale: unable/none/mild/moderate/severe. The apprehension test was performed with the patient in the supine position with the legs relaxed. The examiner stood at the feet of the patient and used their thumbs to exert sufficient lateral force onto the patella to elicit a positive test. The patient's face was observed for the presence of discomfort or apprehension during the maneuver. Positive or negative, the lateral pressure would be discontinued before dislocation of the patella. The amount of apprehension was also graded none, mild, moderate, or severe. The qualitative nature of the patient's apprehension was described using multiple descriptors. Verbal confirmation was selected when the patient informed the examiner that they felt

uncomfortable with the test, but there was no physical reaction. Reflex quadriceps activation was chosen when the quadriceps activated in response to the apprehension test but there was no withdrawal of the leg. Withdraw/squirm was chosen when the patient physically withdrew their leg from the examination. Refuse testing was chosen when the patient would not allow any translation of their patella by firmly co-contracting their thigh muscles or taking other action to avoid assessment. Physiologic was chosen when the patient exhibited or reported sympathetic symptoms such as flushing, sweating, or indicated heart rate changes in response to testing.

To measure the inter-rater reliability, an intraclass correlation coefficient (ICC 2,k) was used for both legs independently for Q-angle measurements. ICC values <0.5, 0.5 to 0.75, 0.75 to 0.9, and >0.90 were considered representative of poor, moderate, good, and excellent reliability, respectively.<sup>9</sup> A Cohen's kappa was used to measure the inter-rater reliability for the J-sign and the apprehension test. Kappa values were interpreted according to the following classification: values ≤ 0 indicating no agreement, 0.01 to 0.20 as none to slight agreement, 0.21 to 0.40 as fair agreement, 0.41 to 0.60 as moderate agreement, 0.61 to 0.80 as substantial agreement, and 0.81 to 1.00 as almost perfect agreement.<sup>10,11</sup> Percent agreement of the raters was also calculated for the inter-rater reliability of the dichotomous classification of the J-sign and apprehension tests. All data were analyzed using SPSS version 27 (IBM, Armonk, NY). The study received ethics approval from the University of Calgary Conjoint Health Research Ethics Board and Child Health Research Office.

## Results

The subjects consisted of 38 patients (11 male, 27 female); age 24.7 (standard deviation 9.4); body mass index 23.6 (standard deviation 2.4), 13 with bilateral symptoms, who presented for initial tertiary orthopaedic surgical consultation for recurrent lateral patellofemoral instability.

Inter-rater reliability for the physical examination tests are presented in [Table 1](#). The mean Q-angle assessed on the right limbs was 9.92° and on the left limbs was 10.58°. The measurement of the Q-angle demonstrated poor inter-rater reliability. The assessment of the J-sign demonstrated fair to moderate reliability, and the evaluation of apprehension demonstrated fair to substantial reliability, with variability evident based on the limb side assessed. When the J-sign was graded on a dichotomous scale as positive or negative, the examiners agreed in 29/38 subjects (76.3%) for the right knees and 26/38 subjects (68.4%) for the left knees. For the apprehension test, the examiners agreed on the positive and negative classifications in 27/38 subjects (71.1%) for the right knees and

**Table 1.** Inter-Rater Reliability of the Physical Examination Test for Patellofemoral Instability

Physical Exam Test (Limb Side)	Level of Agreement	95% Confidence Interval	Agreement Classification <sup>9,10</sup>
Q-Angle (right)	ICC 2, $\kappa$ = 0.44	0.14 - 0.67	Poor
Q-Angle (left)	ICC 2, $\kappa$ = 0.18	0.15 - 0.47	Poor
J-Sign (right)	$\kappa$ = 0.31	0.08 - 0.54	Fair
J-Sign (left)	$\kappa$ = 0.41	0.18 - 0.64	Moderate
Apprehension Test (right)	$\kappa$ = 0.65	0.45 - 0.85	Substantial
Apprehension Test (left)	$\kappa$ = 0.30	0.08 - 0.52	Fair

28/38 subjects (73.7%) for the left knees. These percent agreement results indicate moderate to substantial agreement. Assessment of the quality of apprehension in terms of the physical and verbal reactions of the subjects demonstrated consistency between the 2 examiners as shown in Table 2.

## Discussion

This study has demonstrated that 3 common physical examination tests for recurrent lateral patellofemoral instability demonstrate weak inter-rater reliability. Although the Q-angle, J-sign, and apprehension tests are routinely included in the assessment of patients with recurrent instability, these tests did not demonstrate reliability between examiners for assessments conducted on the same patients with recurrent lateral patellofemoral instability. This study used grading for the apprehension and J-sign tests in addition to a dichotomous positive or negative scale. Unfortunately, these graded assessments did not yield improved reliability results. One potential explanation for the weak reliability may be the lack of reference standards regarding how to perform the test or classify the result. Another consideration is that the tests as described cannot repeatedly measure what they were designed to assess. Importantly, without strong reliability these tests cannot be valid.<sup>12</sup>

Clinical tests form the foundation of the orthopaedic diagnostic process. A test's ability to perform reliably or to measure something consistently is an essential characteristic for application in clinical practice,<sup>13</sup> as well as research. The statistical estimates of reliability used in the current study included correlation, kappa, and percent agreement. Although percent agreement provides for a more instinctive interpretation, the correlation or kappa provide more information because they take into account agreement beyond pure chance.<sup>9,13</sup> Although the

percent agreement for the J-sign and apprehension tests' dichotomous scales (positive or negative) appear strong, the results indicated that assessors disagreed in 24% to 32% of patients. For a physical examination test to be useful in clinical practice or research, the reliability must reach a substantial level of agreement. Lower levels of agreement indicate greater margins of error and a higher likelihood of false-positive and -negative results. The weak reliability results for the physical exam tests in the present study are also disappointing when compared to tests for other knee ligament injuries; for example, the Lachman and Pivot shift tests have demonstrated 89% to 96% and 79% to 88% sensitivity respectively in repeated studies. These results indicate strong reliability that ensures these clinical tests are a key component of diagnosis and treatment planning.<sup>14</sup>

The lack of reliability in this study may be attributed to the differing levels of clinical experience of the examiners; however, the results were consistent with studies that included experts in the field of patellofemoral instability.<sup>15,16</sup> The clinical tests evaluated in this study are common and routine assessment methods that are taught in medical school and allied health programs. In addition, the tests, including the grading, were reviewed in detail between the raters before commencing the study. Although there is some variation in experience between the 2 raters, they were in fact both experts in musculoskeletal assessment. The weak reliability results in the present study indicate that the lack of validity of the clinical test is more likely the cause of the poor reliability.

The Q-angle is a measure of the lateral force vector on the patella that is created by the pull of the quadriceps muscle through the pulley created by the patella to the attachment of the patellar tendon onto the tibial tuberosity. An increased Q-angle reflects an increased lateral displacing force on the patella, assuming that the measure is a true reflection of the pull of the quadriceps on a patella that is reduced in the trochlear groove. The Q-angle has been used as a determinant for the risk of recurrent patellofemoral instability and thereby used to determine the need for a medializing tibial tubercle osteotomy to reduce the lateral force vector. There are many different descriptions of how to measure the Q-angle that vary with regard to the position of the patient, standing, sitting, or supine, and the position of the knee at varying degrees of knee flexion.<sup>17,18</sup> In a

**Table 2.** Number of Qualitative Ratings for Each of the Assessment Categories in Response to the Apprehension Test

	Examiner 1	Examiner 2
Verbal	24	24
Withdraw	6	2
Reflex	5	7
Refused	0	0
Physiological	14	11

systematic review, Smith et al.<sup>4</sup> determined that the inter-rater reliability of a Q-angle assessment ranged from 0.20 to 0.70 and the intra-rater reliability ranged from 0.22 to 0.75. The lack of standardization of the optimal technique to measure the Q-angle has likely contributed to the ongoing disagreement regarding the reliability and validity of the measurement. In this study the measurement of the Q-angle was standardized by performing the test with the patient supine and the knee in 30° of flexion. At this angle, the patella should be engaged in the trochlear groove, thus providing an element of stability to the patella that was reproducible for the measurement between raters. Despite this standardized approach to testing, the ICCs for the Q-angle in this study were poor. One reason for this poor reliability result in contrast with previous studies<sup>19</sup> could be that the present study included only subjects with patellofemoral instability, which has been associated with abnormal femoral pathoanatomy. With the introduction of the tibial tuberosity-trochlear groove measurement, the Q-angle has been largely replaced using the radiographic evaluation to determine the lateral vector acting on the patella.<sup>17,20,21</sup>

The J-sign is an assessment of the tracking pattern of the patella in the femoral trochlea through a full range of knee motion. It is performed by the examiner observing the patient's knee moving between flexion and full extension and noting the presence of lateral displacement of the patella in full extension, and medial translation during early flexion as the patella re-engages into the femoral groove. Although a well-known component of a knee assessment, information on the details of the grading of the J-sign are unclear. Originally, the J-sign was described as a binary test graded as positive or negative, this is a rather simplistic grading for a complex test that is describing a 3-dimensional motion of the patella. Although more complex classifications have been published, these have not gained widespread acceptance or usage, and the reliability of this more complex grading has not been determined.<sup>22</sup> It is unclear whether it is the tracking of the patella or an asymmetry to the other knee that is considered positive. Because asymptomatic knees have some inherent J-type tracking with movement of the patella from lateral to medial with flexion of the knee,<sup>23,24</sup> the threshold to be considered abnormal is unknown. The International Patellofemoral Study Group (IPSG) evaluated the interobserver and intra-observer reliability of the J-sign in patellofemoral experts and found that the intra-observer reliability was 0.28 or fair ( $P < .01$ ) and the interobserver reliability was 0.53 or moderate; however, the sample size in this study was very small.<sup>15</sup> A more recent IPSG study assessed visual tracking of the patella using videos and determined that in two-thirds of cases the experts correctly identified the presence of patellar mal-

racking; however, the experts correctly graded patellar maltracking in only half.<sup>16</sup> In the present study, fair to moderate inter-rater reliability was noted, consistent with the IPSG findings. On the basis of these research findings, the use of J-sign for predictive value or to guide surgical management should not be undertaken until a valid and reliable classification is available.

The apprehension test measures the uneasiness or anxiety that a patient feels when their unstable patella is stressed laterally. Originally used as a binary diagnostic test for patients with lateral patellofemoral instability,<sup>25</sup> the verbal or nonverbal expression of anxiety or fear that the patella may dislocate is considered to constitute a positive test.<sup>3,6,26</sup> Physical reaction to the lateral pressure, such as reflex contraction of the quadriceps to limit lateral subluxation, is also considered a positive test result.<sup>4</sup> Studies have demonstrated good specificity for the apprehension test for lateral patellofemoral instability in 70% to 92% but poor sensitivity in 7% to 37%.<sup>3,27</sup> In the present study, assessment of apprehension on a more robust 5-point Likert scale was variable, with reliability ranging from fair to substantial. Interestingly, the reliability results for the right and left knees were distinct, raising the question about whether the hand or the eye dominance of the rater may contribute to the rating of this test. In addition, the apprehension test is often performed on the less symptomatic knee first, so there may be some learning curve for the patient and their reaction.

Apprehension is a complex examination sign that may contain physical, emotional, and physiological components. This study completed an initial evaluation of the qualitative nature of the apprehension that a patient exhibits with lateral translation of their patella. Further study into the relationship between the qualitative nature of apprehension and psychological components of patellofemoral instability, such as catastrophizing, will be interesting to explore in the future. Clearly, apprehension is more complex than the dichotomized data that have been the standard to date. It may be that the apprehension tests in the literature were inconclusive because of the fact that the test was considered only positive or negative, whereas recording apprehension on a continuous scale for both its quantitative and qualitative nature would be more appropriate.

### Limitations

Limitations of this study include that it was performed in a tertiary orthopaedic surgery clinic on patients with symptomatic recurrent patellofemoral instability and as such does not represent the full spectrum of patients with patellofemoral instability. Consistency of testing technique for each of the 3 assessments is another potential limitation; however, the raters discussed these tests at length during training, and thus the reliability measurements should have been maximized. Despite

this, the inter-rater reliability of the tests was lower than expected. Considering the multifactorial nature of lateral patellofemoral instability, more comprehensive objective measures with robust reference standards and classifications are essential to correctly assess, diagnose, and treat patients.

## Conclusions

In this study of patients with recurrent lateral patellofemoral instability the common physical examination tests, Q-angle, J-sign, and apprehension test, demonstrated weak inter-rater reliability. These results indicate that these tests are not reliable for communication between health care practitioners or as evaluations for clinical research.

## References

1. Fithian DC, Paxton EW, Stone ML, et al. Epidemiology and natural history of acute patellar dislocation. *Am J Sports Med* 2004;32:1114-1121.
2. Hiemstra LA, Kerslake S, Lafave M, Mohtadi NG. 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* 2016;4:2325967116646085.
3. Malanga GA, Andrus S, Nadler SF, McLean J. Physical examination of the knee: A review of the original test description and scientific validity of common orthopedic tests. *Arch Phys Med Rehabil* 2003;84:592-603.
4. Smith TO, Davies L, O'Driscoll ML, Donell ST. An evaluation of the clinical tests and outcome measures used to assess patellar instability. *Knee* 2008;15:255-262.
5. Fredericson M, Yoon KS. Physical examination and patellofemoral pain syndrome. *Am J Phys Med Rehabil* 2006;85:234-243.
6. Lubowitz JH, Bernardini BJ, Reid JB 3rd. Current concepts review: comprehensive physical examination for instability of the knee. *Am J Sports Med* 2008;36:577-594.
7. Bronstein RD, Schaffer JC. Physical examination of knee ligament injuries. *J Am Acad Orthop Surg* 2017;25:280-287.
8. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. *Stat Methods Med Res* 2004;13:251-271.
9. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155-163.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174.
11. McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)* 2012;22:276-282.
12. Portney LG, WMP. *Foundations of Clinical Research: Applications to Practice (ed 3)*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2009.
13. Khan KS, Chien PF. Evaluation of a clinical test. I: assessment of reliability. *BJOG* 2001;108:562-567.
14. Leblanc MC, Kowalczyk M, Andruszkiewicz N, et al. Diagnostic accuracy of physical examination for anterior knee instability: a systematic review. *Knee Surg Sports Traumatol Arthrosc* 2015;23:2805-2813.
15. Smith TO, Clark A, Neda S, et al. The intra- and inter-observer reliability of the physical examination methods used to assess patients with patellofemoral joint instability. *Knee* 2011.
16. Best MJ, Tanaka MJ, Demehri S, Cosgarea AJ. Accuracy and reliability of the visual assessment of patellar tracking. *Am J Sports Med* 2020;48:370-375.
17. Freedman BR, Brindl TJ, Sheehan FT. Re-evaluating the functional implications of the Q-angle and its relationship to in-vivo patellofemoral kinematics. *Clin Biomech (Bristol, Avon)* 2014;29:1139-1145.
18. Smith TO, Hunt NJ, Donell ST. The reliability and validity of the Q-angle: A systematic review. *Knee Surg Sports Traumatol Arthrosc* 2008;16:1068-1079.
19. Mendonca Lde M, Ocarino JM, Bittencourt NF, Santos TR, Barreto RA, Fonseca ST. Normative data of frontal plane patellar alignment in athletes. *Phys Ther Sport* 2015;16:148-153.
20. Dejour H, Walch G, Nove-Josserand L, Guier C. Factors of patellar instability: An anatomic radiographic study. *Knee Surg Sports Traumatol Arthrosc* 1994;2:19-26.
21. Goutallier D, Bernageau J, Lecudonnec B. [The measurement of the tibial tuberosity. Patella groove distanced technique and results (author's transl)]. *Rev Chir Orthop Reparatrice Appar Mot* 1978;64:423-428.
22. Donell ST, Shepherd K, Ali K, McNamara I. The inferomedial patellar protuberance and medial patellar ossicle in patellar instability. *Knee Surg Sports Traumatol Arthrosc* 2017;25:2682-2687.
23. Amis AA, Senavongse W, Bull AM. Patellofemoral kinematics during knee flexion-extension: an in vitro study. *J Orthop Res* 2006;24:2201-2211.
24. Kujala UM, Osterman K, Kormanen M, Nelimarkka O, Hurme M, Taimela S. Patellofemoral relationships in recurrent patellar dislocation. *J Bone Joint Surg Br* 1989;71:788-792.
25. Fairbank HA. Internal derangement of the knee in children and adolescents: (Section of Orthopaedics). *Proc R Soc Med* 1937;30:427-432.
26. Hughston JC. Subluxation of the patella. *J Bone Joint Surg Am* 1968;50:1003-1026.
27. Haim A, Yaniv M, Dekel S, Amir H. Patellofemoral pain syndrome: Validity of clinical and radiological features. *Clin Orthop Relat Res* 2006;451:223-228.