

Original Research

How Reliable is a J-sign Severity Scale When Assessing Lateral Patellar Instability?

Oksana Klimenko, MD¹; Ted C. Sousa, MD²; Ryan Baker, MD²; Jacob Carl, MD²; Shelley Mader, PT²; Kristopher Holden, MPT²; Mark L. McMulkin, PhD²

¹Elson S. Floyd College of Medicine at Washington State University, Spokane, WA; ²Shriners Children's–Spokane, Spokane, WA

Correspondence: Ted C. Sousa, MD, Medical Staff, Shriners Children's–Spokane, 911 W. 5th, Spokane, WA 99204.
E-mail: tsousa@shrinenet.org

Received: December 9, 2023; Accepted: May 20, 2023; Published: August 1, 2023

DOI: 10.55275/JPOSNA-2023-630

Abstract

Background: Patellar instability is a common cause of anterior knee pain and can limit function and sports participation. To help assess patellar instability, the clinical J-sign test consists of observing the patella translate laterally in the shape of an inverted J over the anterolateral femur proximal to the trochlear groove during active knee extension. Only positive or negative categorization of the J-sign test has typically been used without rating the severity. The purpose of this study was to assess the inter- and intra-observer reliability of a grading/severity scale of the J-sign test.

Methods: A scale for J-sign severity was utilized as follows: grade 0: unable to complete J-sign due to pain or apprehension; grade 1: ≤ 1 quadrant of translation; grade 2: >1 quadrant of translation; grade 3: >2 quadrants of translation; grade 4: complete patellofemoral dislocation. This retrospective cross-sectional study assessed J-sign ratings (0 to 4) from videos of patients undergoing evaluation for patellar instability. Six healthcare professionals rated the severity of the J-sign using the proposed scale, two different times, for all knees presented in random order. Inter- and intra-observer reliability were calculated using a Fleiss Kappa, κ .

Results: Forty-four patients (87 knees) ages 10-18 were included in this study. Both knees were rated, including unaffected knees to serve as a control. The proposed standardized grading scale for the J-sign had fair agreement for inter-observer reliability, $\kappa = 0.31$, and moderate agreement for intra-observer reliability, $\kappa = 0.58$.

Conclusion: The proposed scale for determining J-sign severity yielded fair inter-observer reliability and moderate intra-observer reliability, similar to the Kappa scores evaluating only the presence or absence of the J-sign. Further study into developing a standardized scale for J-sign severity grading might improve clinical descriptors of the test and expand on other factors, including clarity of knee extension ability, video standardization, and training materials.

Level of Evidence: Level III, retrospective cross-sectional study

Key Concepts

- Inter-observer agreement on a 0 to 4 rating scale of J-sign for lateral patellofemoral maltracking was fair while intra-observer agreement was moderate.
- The four-quadrant scale of patellofemoral instability may not reproducibly assess lateral patellar maltracking.
- Further development of a revised grading scale for lateral patellar maltracking is warranted.

Introduction

Patellar instability is a common cause of anterior knee pain limiting function and sports participation. A clinical J-sign test can be employed to assess the presence of patellar mal-tracking and lateral instability. A positive J-sign occurs when the patella laterally translates over the anterolateral femur proximal to the trochlear groove during active knee extension.¹ As the knee is fully extended from 30-degrees flexion, the J-sign is considered positive if the patella tracks in an inverted “J” shape. Overall, the J-sign test is used to determine patella hypermobility and possible instability. It is a common clinical test that is a simple and inexpensive method to help determine need for a more complex workup since patellar instability is associated with osseous and/or soft tissue abnormalities. Positive or negative categorization is often currently used in clinical practice, with categorical severity ratings having been previously studied in literature.²⁻⁶ A positive clinical J-sign along with symptoms of instability may indicate the need for intervention. However, the simple presence or absence of a clinical J-sign does not specify the severity of lateral instability limiting management and communication. Less severe patellar instability might guide treatment toward more conservative therapy, while more severe instability might tend toward less conservative treatment.

An objective classification system for the J-sign test has been assessed for patients presenting with repetitive patellar instability. The correlation of a J-sign severity (3-category scale) to interventional outcomes in patients with repetitive patellar instability has been reported.² The J-sign grading scale was found to have an inter- and intra-observer reliability of $\kappa = 0.78$ and 0.84 including

only participants with a prior positive J-sign.² Another study tested the inter- and intra-observer reliability of determining the presence or absence of a J-sign, $\kappa = 0.53$ and 0.28 , respectively.³ In a study with orthopaedic surgeons comparing the qualitative and four-quadrant J-sign scale, the inter-observer reliability for determining solely the presence of a clinical J-sign, $\kappa = 0.45$, was not much higher than the inter-observer reliability of the quadrant scale, $\kappa = 0.42$.⁶ A more complete inter- and intra-observer reliability analysis of two J-sign severity scales quadrant classification (grade 1 to 4) and Donell classification (8 categories) has been reported.⁴ Inter-rater reliability between orthopaedic specialists was found to be 0.51 and 0.49 for the quadrant and Donell classifications; intra-rater reliability was 0.79 and 0.72 , respectively. These are moderate to substantial agreements. However, all assessors were aware that only participants with recurrent patellar instability were included viewing both knees in the same video to assess symmetry (43% had unilateral involvement). It would be helpful to follow-up this study/quadrant classification with a scale and assessments, including J-sign tests that could not be completed due to apprehension/pain. It is also important to utilize a variety of healthcare professionals, to determine the reliability beyond a single discipline and increase the generalizability of study findings.

The purpose of this retrospective study was to assess the inter- and intra-observer reliability of a grading scale of J-sign severity. Creating a reproducible severity grading scale will allow future studies to assess the relevance to patient prognosis and

outcomes. This scale will include a range of J-sign severity, including “typical” values with minimal lateral translation to severe lateral displacement. A separate category to include patients with lateral instability unable to perform the exam due to pain or apprehension was also examined. It was hypothesized that utilizing a four-quadrant dislocation system, with a rating for inability to complete test, will lead to a reproducible severity scale for the J-sign test.

Materials and Methods

Study Design

Patients that underwent a complete gait study between January 2016 and October 2020 with a diagnosis of patellar instability (either unilateral or bilateral) and had a J-sign test evaluation completed with video recording serving as documentation were considered for inclusion. After institutional IRB approval, we retrospectively identified these patients’ videos. Videos of the patients’ J-sign assessment with bilateral or unilateral knee involvement were included, as the unaffected knee could serve as a control. Patients were excluded if they had previous surgeries or identifiable scarring/markings on their knee, as this would affect the blinding process. Patients were also excluded if they had a flexion contracture greater than 10 degrees, which could affect the ability to reach terminal extension.

J-sign Evaluation and Grading Scale

J-sign severity was determined by lateral patellar translation over the anterolateral trochlea when the knee was taken to terminal extension from 30 degrees

of flexion. Previously proposed scales^{2,4-6} were modified to include those unable to complete the test due to pain or apprehension. J-sign severity was graded from 0 to 4 determined with a “quadrants of translation” classification scale (grade 0: unable to complete J-sign due to pain or apprehension; grade 1: <1 quadrant of translation; grade 2: >1 quadrant of motion; grade 3: >2 quadrants of motion; grade 4: complete patellofemoral dislocation) (Table 1). A visual depiction the J-sign severity scale was provided to the raters (Figure 1).

Videos of patients undergoing evaluation for patellar instability via the J-sign test in the Motion Analysis Center of a Children’s Hospital were used to evaluate the severity. The patella was first outlined with a marking pen. Then a video, capturing only the legs, was taken from a frontal view of the anterior knee (See Video). Each video assessed one limb at a time. Seated on the edge of a table with legs hanging over the side, the patient slowly extended their knee to terminal extension. Still pictures of the starting and ending position of the J-Sign test are shown in Figure 2. Six healthcare professionals (two pediatric orthopaedic surgeons, one sports medicine pediatrician, two physical therapists, one medical student) were blinded to the participants’ videos and assessed the severity of the J-sign using the proposed scale for two separate trials separated by at least 3 weeks. Videos of knees from the same patient were not shown together or consecutively so raters did not have knowledge if an individual video was an affected knee.

Table 1. J-sign Severity Classification

Grade	Classification
0	Unable to complete J-sign due to pain or apprehension
1	Mild: gentle or normal. ≤ 1 quadrant of translation
2	Moderate: >1 and up to 2 quadrants of translation
3	Severe: >2 quadrants of translation
4	Habitual dislocation in extension: complete patellofemoral dislocation with knee extension

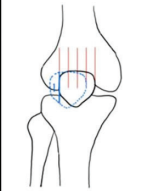
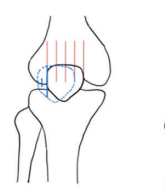
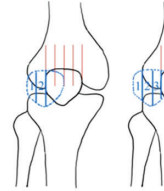
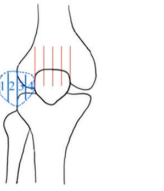
Scale Rating & Description	0: Test not completed. Knee not fully extended, likely due to apprehension	1: Mild. Gentle or normal. ≤ 1 quadrant of lateral translation	2: Moderate >1 & ≤ 2 quadrants of lateral translation	3: Severe. >2 & < 4 quadrants of lateral translation	4: Habitual dislocation in extension. Patella dislocates completely in extension with each knee cycle of flexion/extension
Depiction of Quadrants of Lateral Translation		 1 Quadrant Lateral Translation	 1 Quadrant Lateral Translation 2 Quadrants Lateral Translation	 2 Quadrants Lateral Translation 3 Quadrants Lateral Translation	 4 Quadrants Lateral Translation

Figure 1. Visual depiction of J-sign severity scale.

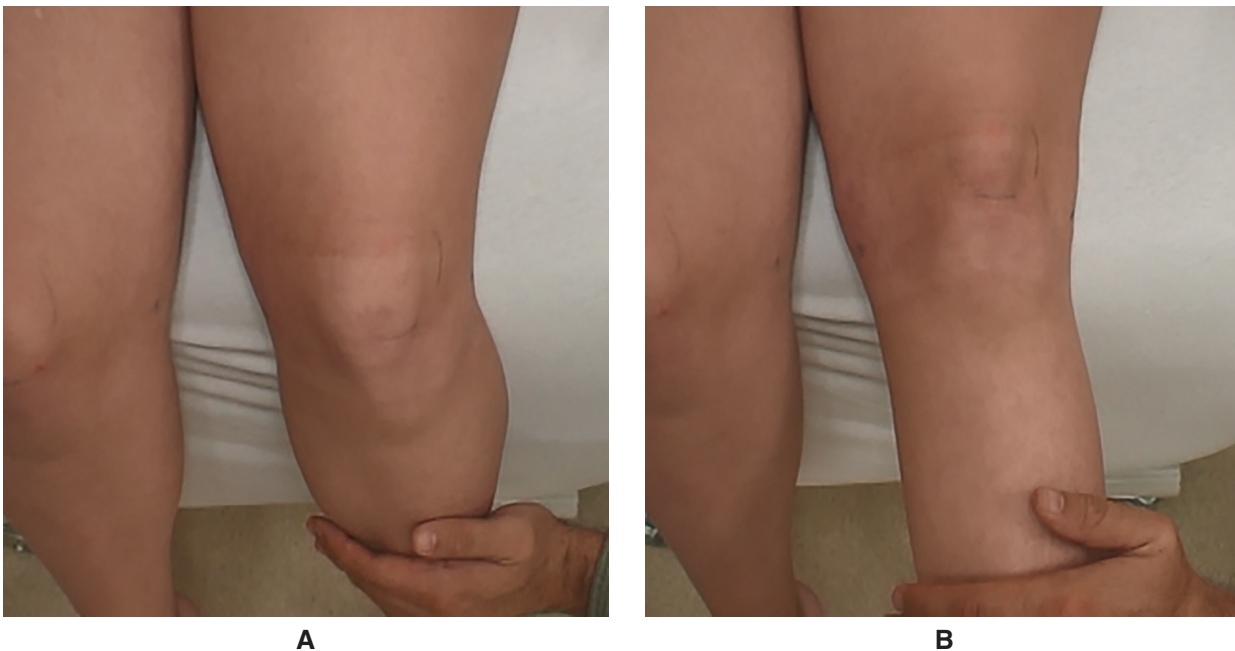


Figure 2. Example still pictures of participant with knee flexed (A) and knee extended (B). Pictures are intended to show how the test was completed. Videos were used for rating scale.

Statistical Analysis

Inter- and intra-rater reliability was determined using the Fleiss' Kappa (κ), as the ratings were considered categorical data requiring complete agreement between ratings. Analyses were completed using R software, version 4.0.3 (Foundation for Statistical Computing, Vienna, Austria). The following parameters dictated the agreeability using the Fleiss' Kappa: $\kappa < 0$: poor agreement; 0.01-0.20: slight agreement; 0.21-0.40: fair agreement; 0.41-0.60: moderate agreement; 0.61-0.80:

substantial agreement; and 0.81 -1.00 almost perfect agreement.⁷

Results

A total of 48 patients were identified that met the criteria of diagnosis of patellar instability with an available video of their J-sign test. Four patients were excluded due to previous surgery and the presence of visible scars around the knee. One patient did not have a video of the unaffected knee; however, the video for the involved

Table 2. Participant Demographics

Characteristic	Result
Number participants/Number knees (n)	44/87
Number symptomatic/Number asymptomatic knees	63/24
Age (years), mean (standard deviation), range	14.4 (2.1), range 10.2-17.9
Gender female/male (n)	33/11

knee was included. Nineteen patients contributed bilaterally affected knees. The unaffected knee of the remaining 24 patients with unilateral involvement were considered control ratings. Therefore, a total of 87 knees were evaluated by six raters. The characteristics of the patients included in the study are shown in Table 2. Female to male ratio of patients was 3 to 1.

The Fleiss Kappa for inter-rater reliability was found to be $\kappa = 0.29$ and $\kappa = 0.32$ for the first and second set of ratings, respectively (raters = 6, n = 87 ratings). Overall, the ratings were found to be in “fair agreement” with the kappa values for both trials between 0.21 to 0.40. The intra-rater reliability was calculated to be of “moderated agreement” with $\kappa = 0.58$ (repetition = 2, n = 522 ratings 6 raters X 87 videos,).

Discussion

The purpose of this study was to assess inter- and intra-rater reliability of a J-sign grading scale. The studied scale showed fair agreement between evaluators and moderate agreement of multiple ratings done by the same evaluator. This study differed from previous ones in that it included a category for inability to complete maneuver due to apprehension, a more diverse evaluator group utilizing healthcare professionals in different specialties and resulted in a worse inter-observer reliability than prior studies.

Previous studies of different J-sign rating scales have reported inter-observer κ of 0.51 to 0.78 and intra-rater reliability κ of 0.28 to 0.84.^{2-4,6} These studies also utilized the quadrant system. The current study found a somewhat lower inter-rater reliability ($\kappa = 0.29$)

and an intra-rater reliability within the same range. Differences in the studies included sample size of knees to be rated (n = 10 to 30, compared to current study n = 87), number of raters (raters = 2 to 30, compared to current study raters = 6), and severity scale that was used. Unlike previously explored scales, the current study included the lack of a J-sign due to apprehension to complete the evaluation. This difference did not lead to improved inter- and intra-rater reliabilities. These results confirm that further development and refinement of a J-sign severity scale is needed in order to reach clinical utility. When tested with a variety of healthcare professionals, the inter-rater reliability decreased from previous studies that utilized mainly orthopaedic surgeons as raters.

Previous studies also used orthopaedic surgeons as the majority to evaluate the inter- and intra-observer reliabilities of a proposed J-sign severity scale.²⁻⁴ This study used a more diverse evaluator panel including two pediatric orthopaedic surgeons, one sports medicine pediatrician, two physical therapists, one medical student. This allowed for the assessment of the scale’s reproducibility and applicability of use among a wide array of healthcare professions and specialties. This could also contribute to the lower inter-observer reliability found in this study compared to previous ones. Not only does the four-quadrant scale have a low reproducibility when studied solely amongst orthopaedic surgeons,⁴ but this study showed a further decrease in reproducibility when involving other healthcare professionals. This shows a lower utility for it in a variety of healthcare settings, from orthopaedic surgery, physical therapy, primary care, and medical education.

Other physical exam tests about the knee have been shown to have the following Fleiss Kappa values for the inter-observer reliability: McMurray: 0.21, Anterior Drawer: 0.65, Lachman: 0.55, Pivot-shift: 0.57, Lever sign 0.88.^{8,9} With this study's proposed severity scale for the J-sign yielding an inter-observer reliability of $\kappa = 0.31$, its reproducibility was somewhat lower than several other orthopaedic physical exam tests. This indicates the potential for further development and exploration of a novel rating scale for the J-sign or determining other methods to evaluate patellar instability.

Several factors contributed to the limitations of this study. The videos varied slightly by angles of capture, room lighting, camera movement, and completion by different evaluators. The evaluators had somewhat different hand placement to support the patient's thigh. Future studies could consider standardization of the video documenting approach as well as either marking the four quadrants prior to ratings or using less quadrants to minimize variation in perception and interpretation. Additionally, a sagittal view of the knee extension attempt could be useful to assess the degree of extension achieved. The current approach to assessing the J-sign test matched a more realistic clinical implementation. Displacement of the patella was the only factor rated by the scale used in the current study. However, speed of patellar displacement might have influenced the ratings. For instance, the appearance of a rapidly subluxating patella might have led to a higher severity rating. In addition, body mass index might have impacted the ability to discern the patella. Repeatability by body mass index ranges were not considered. The ratings used in the present study were analyzed as categorical data, utilizing the unweighted Fleiss Kappa. The main reason for this was inclusion of the category of "unable to complete the test due to apprehension of subluxation (rating of 0)." If the inability to complete the test at all were to be considered more severe than a complete four-quadrant subluxation, then ordinal data instead of categorical data could be assumed. If the ratings were instead considered interval data and utilized

an ICC, the degree of agreement between raters would be greater.

This assessment of the four-quadrant system continues to yield lower reproducibility than those of the previous studies. A previous study indicated fair agreement in the four-quadrant J-sign, not much different than the κ of qualitative scale of indicating the presence or absence of the J-sign.⁶ Furthermore, the current study's lower $\kappa = 0.31$, in the setting of ratings from various healthcare fields, may point to the four-quadrant scale's greater lack of utility outside the orthopaedic surgery field. While an argument can still be made that creating a reproducible J-sign severity scale may aid in interventional algorithms for patellar instability, a scale with higher reproducibility needs to be developed to determine that correlation.

In conclusion, this study presents data for the development of a quantitative scale for the J-sign. The reliability measures were fair to moderate. Given that patellar instability is a common cause of pain and dysfunction, a quantification system of the physical exam findings of lateral translation, the J-sign, could be useful in the evaluation of patellar maltracking and improve communication between providers assessing patients with patellar instability. However, future studies must continue to refine and improve the reliability of the J-sign when assessing patients for patellar instability; otherwise, it may be wise to explore different scales outside the four-quadrant scale.

Additional Links

- [Video illustrating J-sign testing procedures](#)

Disclaimer

No funding was received. The authors report no conflicts of interest related to this manuscript.

References

1. Beckert MW, Albright JC, Zavala J, et al. Clinical accuracy of J-sign measurement compared to magnetic resonance imaging. *Iowa Orthop J.* 2016;36:94-97.
2. Zhang Z, Zhang H, Song G, et al. A high-grade J sign is more likely to yield higher postoperative patellar laxity and residual maltracking in

- patients with recurrent patellar dislocation treated with derotational distal femoral osteotomy. *Am J Sports Med.* 2020;48(1):117-127.
3. 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.* 2012;19(4):404-410.
 4. Hiemstra LA, Sheehan B, Sasyniuk TM, et al. Inter-rater reliability of the classification of the J-sign is inadequate among experts. *Clin J Sport Med.* 2022;32(5):480-485.
 5. Franciozi CE, Ambra LF, Albertoni LJB, et al. Anteromedial tibial tubercle osteotomy improves results of medial patellofemoral ligament reconstruction for recurrent patellar instability in patients with tibial tuberosity-trochlear groove distance of 17 to 20mm. *Arthroscopy.* 2019;35(2):566-574.
 6. Best MJ, Tanaka MJ, Demehri S, et al. Accuracy and reliability of the visual assessment of patellar tracking. *Am J Sports Med.* 2020;48(2):370-375.
 7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.
 8. Galli M, Ciriello V, Menghi A, et al. Joint line tenderness and McMurray tests for the detection of meniscal lesions: what is their real diagnostic value? *Arch Phys Med Rehabil.* 2013;94(6):1126-1131.
 9. Bilgin E, Turgut A, Hancioğlu S, et al. The influence of anesthesia-body mass index and chronicity of the injury on the reliability of diagnostic tests for anterior cruciate ligament rupture. *J Exerc Rehabil.* 2021;17(6):428-434.