BJO

KNEE

A comparative analysis of interobserver reliability and intraobserver reproducibility of the Oswestry-Bristol Classification and the Dejour Classification for trochlear dysplasia of the knee

K. D. Roy,
P. Joshi,
I. Ali,
P. M. Shenoy,
I. Malek,
D. Barlow,
A. Syed,
Y. Joshi

From Wrexham Maelor Hospital, Wrexham, UK

Aims

Classifying trochlear dysplasia (TD) is useful to determine the treatment options for patients suffering from patellofemoral instability (PFI). There is no consensus on which classification system is more reliable and reproducible for the purpose of guiding clinicians' management of PFI. There are also concerns about the validity of the Dejour Classification (DJC), which is the most widely used classification for TD, having only a fair reliability score. The Oswestry-Bristol Classification (OBC) is a recently proposed system of classification of TD, and the authors report a fair-to-good interobserver agreement and good-to-excellent intraobserver agreement in the assessment of TD. The aim of this study was to compare the reliability and reproducibility of these two classifications.

Methods

In all, six assessors (four consultants and two registrars) independently evaluated 100 axial MRIs of the patellofemoral joint (PFJ) for TD and classified them according to OBC and DJC. These assessments were again repeated by all raters after four weeks. The inter- and intraobserver reliability scores were calculated using Cohen's kappa and Cronbach's α.

Results

Both classifications showed good to excellent interobserver reliability with high α scores. The OBC classification showed a substantial intraobserver agreement (mean kappa 0.628; $p < 0.005$) whereas the DJC showed a moderate agreement (mean kappa 0.572; $p < 0.005$). There was no significant difference in the kappa values when comparing the assessments by consultants with those by registrars, in either classification system.

Conclusion

This large study from a non-founding institute shows both classification systems to be reliable for classifying TD based on axial MRIs of the PFJ, with the simple-to-use OBC having a higher intraobserver reliability score than that of the DJC.

Cite this article: *Bone Jt Open* 2023;4-7:532–538.

Keywords: Knee, Trochlear dysplasia, Reliability, Dejour Classification, Oswestry-Bristol Classification, Patellofemoral dysplasia, Interobserver, Intraobserver

Introduction

Patellofemoral instability (PFI) is a complex condition and its aetiopathogenesis is thought to be multifactorial in most cases. The limb alignment, knee anatomy (including both patellar and trochlear shape), and static and dynamic constraints help to maintain the stability of the patellofemoral joint (PFJ).[1] Trochlear dysplasia (TD) is thought to be one of the more prevalent and important factor

Correspondence should be sent to Kunal Dwijen Roy; email: kunaldroy@gmail.com

doi: 10.1302/2633-1462.47.BJO-2023-0048

Bone Jt Open 2023;4-7:532–538.

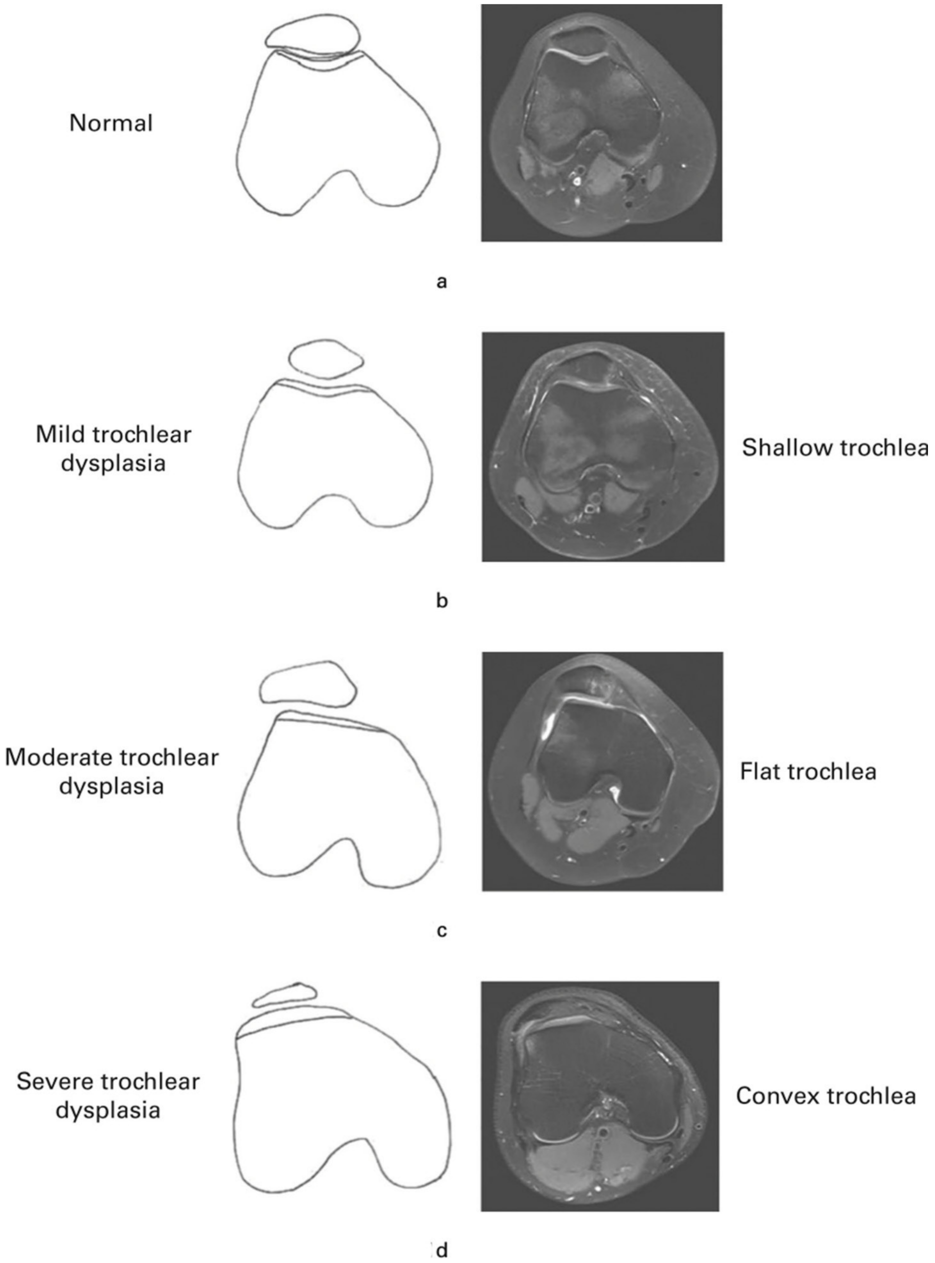


Fig. 1

Oswestry-Bristol Classification (illustration and MRI reproduced with permission from Sharma et al).[6]

contributing to PFI, with up to 96% of patients with a history of patellofemoral dislocation having evidence of trochlear dysplasia.[2,3]

The classification of TD is important to help decide the treatment options for PFI. Though there is no consensus on which classification system is more reliable and reproducible for this purpose, the Dejour Classification (DJC) is widely used around the world.[3,4] However, there are concerns about its validity and reliability scores in the literature.[5]

The Oswestry-Bristol Classification (OBC) is a recently proposed system of classification of TD and the authors report a fair-to-good interobserver agreement and good-to-excellent intraobserver agreement.[6] This study aimed to compare the reliability and reproducibility of these two classifications.

Methods

This study was conducted at a large district general hospital in the UK. The hospital database was searched for MRI scans.

MRI scans of the knee performed in the past three years were used to identify patients with TD. We used the search terms ‘trochlear dysplasia’ and ‘shallow’ in the

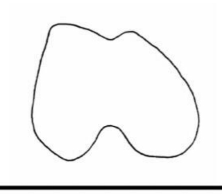
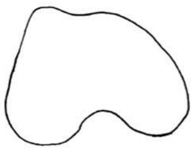
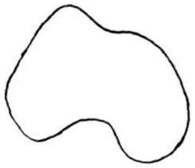
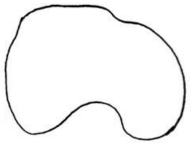
Dejour Type A	Shallow Trochlea >145 degrees	
Dejour Type B	Flat Trochlea	
Dejour Type C	Lateral convexity, Medial hypoplasia	
Dejour Type D	Cliff	

Fig. 2

Representation of Dejour Classification.[7]

MRI scan report from the picture archiving and communication system (SYNAPSE; Fujifilm, Japan). These MRIs were reviewed by the lead investigator (KDR) and senior author (YJ) to select 95 MRI scans conforming to the inclusion and exclusion criteria.

Scans of patients over the age of 16 and under 65 with visible TD on the scan were included. Those with signs of previous patellofemoral surgery, patellofemoral arthritis, or suboptimal images were excluded. Also included in the cohort of scan images were five normal knee MRI scans which did not display any features of TD.

A total of six assessors (four orthopaedic consultants and two orthopaedic specialty registrars) independently evaluated the 100 knee axial MRIs of the PFJ for signs of TD. They were asked to classify these according to OBC and DJC in two separate sittings, with at least 24 hours between the sittings. The image sequence was randomized for each assessment sitting, and the assessors were blinded to the sequence to eliminate bias. These assessments were again repeated by all raters after four weeks. They were provided with detailed illustrations of the OBC and DJC during each sitting, as a reference guide (Figures 1 and 2).

We also performed the following subanalysis: 1) interobserver variability among consultants compared with registrars; and 2) interobserver variability among a soft-tissue knee surgeon (YJ) and non-specialist orthopaedic doctors.

Table I. Agreement measures according to Landis and Koch.[8]

Kappa statistic	Strength of agreement
< 0.00	Poor
0.00 to 0.20	Slight
0.21 to 0.40	Fair
0.41 to 0.60	Moderate
0.61 to 0.80	Substantial
0.81 to 1.00	Almost perfect

These two subanalyses were performed to determine whether seniority or specialization affected the use of these classifications.

Statistical analysis. The statistical analysis, including calculations of magnitude of agreement (observed agreement), Cohen's kappa statistics for intraobserver reproducibility, and Cronbach's α values for interobserver agreement, was calculated using SPSS 26.0 (IBM, USA).

The observed agreement is defined as the proportion of cases where the two observers (or same observer on two occasions) agree. Given that there is a mathematical probability of two observers giving the same response for any given case (chance agreement), the kappa coefficient is the observed agreement over and above that due to chance.

Table II. Intraobserver scores for reading 1 and reading 2 for the Oswestry-Bristol Classification.

Assessor	Kappa	p-value
Assessor 1	0.422	< 0.001
Assessor 2	0.673	< 0.001
Assessor 3	0.639	< 0.001
Assessor 4	0.581	< 0.001
Assessor 5	0.698	< 0.001
Assessor 6	0.755	< 0.001
Mean	0.628	

Cohen’s kappa was run to determine if there is agreement between the scores for reading 1 and reading 2 of the Oswestry-Bristol Classification. There was substantial agreement between these two readings (p < 0.0005).

Table III. Intraobserver scores for reading 1 and reading 2 for the Dejour Classification.

Assessor	Kappa	p-value
Assessor 1	0.625	< 0.001
Assessor 2	0.585	< 0.001
Assessor 3	0.534	< 0.001
Assessor 4	0.631	< 0.001
Assessor 5	0.610	< 0.001
Assessor 6	0.449	< 0.001
Mean	0.572	

Cohen’s kappa was run to determine if there is agreement between the scores of raters for reading 1 and reading 2 of the Dejour Classification for trochlear dysplasia. There was moderate agreement between these two readings (p < 0.0005).

Table IV. Comparison of intraobserver scores of consultants and registrars for the Oswestry-Bristol Classification.

Assessor	Mean kappa value
Consultants (Assessors 1 to 4)	0.578
Registrars (Assessors 5 and 6)	0.726

Table V. Comparison of intraobserver scores of consultants and registrars for the Dejour Classification.

Assessor	Mean kappa value
Consultants (Assessors 1 to 4)	0.593
Registrars (Assessors 5 and 6)	0.529

Table VI. Comparison of intraobserver scores of knee specialist and other consultants for the Oswestry-Bristol Classification.

Assessor	Mean kappa value
Knee specialist	0.673
Other consultants	0.599

Table VII. Comparison of intraobserver scores of knee specialist and other consultants for the Dejour Classification.

Assessor	Mean kappa value
Knee specialist	0.585
Other consultants	0.578

Landis and Koch[8] have suggested that kappa values falling in different ranges imply different degrees of agreement (Table I). A kappa coefficient value < 0.00, i.e. a negative kappa, suggests poor agreement, 0.00 to 0.20 slight agreement, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial, and 0.81 to 1.00 almost perfect or excellent agreement. A p-value < 0.005 was considered significant.

Results

Scans belonging to 94 patients were included in this study (six bilateral scans). The mean age of patients whose MRI scans were included was 29 years (17 to 58). Of these, 63 were female and 31 male.

Based on the responses of all the assessors for OBC, 5% of the MRI scans were classified as ‘normal’, 17% as ‘mild’, 49.5% as ‘moderate’, and 28.5% as ‘severely dysplastic’. During the assessment of the MRI scans according to the DJC ; 4.75% were classified as ‘normal’, 9.75% as Grade A, 26.33% as Grade B, 33.5% as Grade C, and 25.66% as Grade D.

We calculated the Cohen’s kappa value to determine the agreement between the two readings of each assessor for both classifications (Tables II and III). The OBC showed ‘substantial agreement’ with a kappa value of 0.628. On the other hand, the DJC only showed ‘moderate agreement’ with a kappa value of 0.572.

Both classifications showed excellent interobserver agreement during each of the readings. However, the OBC had a marginally higher value of Cronbach’s α at both readings (0.930, 0.945) compared with the readings for the DJC (0.925, 0.930).

The mean kappa of the registrars was higher than that of the consultants while using the OBC (0.726 vs 0.578), whereas the consultants had a higher average kappa while using the DJC (0.593 vs 0.529), as shown in Tables IV and V. Neither of these findings reached statistical significance.

The kappa value for the readings by the knee specialist was higher than that of the non-specialist consultants for both the DJC and OBC (Table VI and Table VII), however, again neither of these were statistically significant values.

Discussion

Treatment of PFJ instability depends on studying individual knee anatomy and assessing the various stabilizing factors through appropriate investigations and analysis of any morphological abnormalities found. Additionally, the cartilaginous morphology of the knee in patients with TD differs from the bony morphology as shown by van Huyssteen et al[9] in their MRI-based study. This difference needs to be considered when planning surgical options for these patients.

Various measurements have been described to study the degree and type of TD such as trochlear bump, trochlear depth, crossing sign, and lateral trochlear inclination.[1,2]

In their seminal work on PFI, Dejour et al[10,11] proposed a classification system for TD, initially identifying three types of dysplasia on lateral radiograph and cross-sectional CT images. This was later modified to have four grades (Type A to Type D).[7] This classification has been widely used by orthopaedic surgeons to measure the severity of TD.[12–14] Many of these descriptions are based on a lateral knee radiograph and a strict lateral view is of considerable importance. Any deviation from this can lead to erroneous interpretations and classification of TD.[4,15]

Globally, in current orthopaedic practice, the use of MRI has superseded other methods for the evaluation of TD and PFI.[15–17] Lippacher et al[3] adapted the original DJC for CT scan into a four-grade MRI classification for TD. However, in a comparative analysis of the radiological and MRI-based DJC, the founding authors noted that the MRI-based four-grade DJC for TD only showed fair intraobserver and interobserver agreement. Though this was better than that for lateral radiographs, it was deemed unsatisfactory by the authors. They attributed these unsatisfactory results to the complex geometry of the trochlea and differences in interpretation.

We have had a similar experience in our study, with the raters observing that it was quite difficult to distinguish between a Type B and Type C trochlea. Features of hypoplasia can be underinterpreted as a descending/flat trochlea and vice versa and, similarly, a slight vertical element on a trochlea with other features of a Type C trochlea can be classified as Type D. This also has a bearing on the treatment and surgical plan of management of these patients, as the operative recommendations for each grade is different.

Though the DJC may well be the most widely used classification for TD, current evidence shows that it has a highly variable reliability and is not particularly useful in clinical settings and when directing treatment.[5,18,19] Authors have also questioned its accuracy in the grading of TD and its severity.[5] As mentioned earlier, the founding group themselves found low intraobserver and interobserver reliability scores for the MRI-based four-grade DJC.[3,5] In this study, four surgeons independently evaluated lateral radiographs, and then axial MRI scans of the knee, on two occasions four weeks apart. The intraobserver reliability for the MRI scans ranged from 32% to 74% and the interobserver reliability was only 28% to 60%.

In their study on classification systems and radiological measurements for TD in skeletally immature patients, Stepanovich et al[18] had similar findings of low intraobserver and interobserver reliability for the DJC, with the kappa values being 0.596 and 0.687, respectively.[5,18]

In their analysis, Sharma et al[6] found the DJC to have poor interobserver reliability on calculating kappa scores. This was found to be fair to good when this was expressed in terms of the S-statistic. The intraobserver reliability varied from kappa values of 0.27 to 0.78, with a mean value of 0.52 signifying fair to good agreement.

In the present study, we have found that the intraobserver reliability for the DJC ranged from kappa coefficient values of 0.449 to 0.631, with a mean kappa of 0.572 signifying moderate agreement. We used Cronbach's α score to compare and quantify the interobserver reliability, as this gives a better picture of internal consistency when multiple assessors have been employed.[20,21] The α scores for the DJC in the first and second rounds were 0.925 and 0.93, respectively, indicating excellent interobserver reliability. This is in contrast with other studies in the literature that have shown a poor/fair interobserver score for the DJC. We believe this could be due to the comparatively high number of cases/MRIs reviewed by the assessors in this study, which has consequently led to an improvement in the reliability scores by reducing any confounding that may occur due to a smaller sample.

Various quantitative parameters have also been described in the literature to assess and grade the severity of TD. The lateral trochlear inclination (LTI) angle was described in a French study in 2000 as a sensitive and accurate representation of the degree of TD.[4,22] It was also found to be highly reproducible, with an excellent ability to distinguish between high- and low-grade dysplasia using 11° value for LTI as the threshold margin.[22] Pfirrmann et al[23] measured the trochlear depth, facet asymmetry, condyle asymmetry, and lateralization of patella on axial MRI scans of the knee. They found these parameters to be reliable for diagnosing TD on axial MRIs when measurements were made 3 cm above the joint space. Measurements and ratios of the height of medial and lateral trochlea to the width of trochlea have also been proposed to identify the site of pathology and severity of TD by Biedert and Bachmann.[24] Using these parameters, they found most cases of TD to have a pathology in the medial or central part of the femoral condyle. A comparative MRI-based analysis of these quantitative measurements by Nelitz et al[4] in 2014 shows they have only limited value in the assessment of TD. In addition, they observed poor correlation between these measurements and the DJC.

The recently described OBC for TD is purportedly easier to use, with better reliability than the DJC.[6] The authors claim to have noticed that the DJC was not useful in guiding treatment of patients with TD, and hence developed this classification.[6,25] The OBC grades TD based on axial MRIs as normal, mild, moderate, and severe dysplasia, which correspond to normal, shallow, flat, and convex trochlea. The main criterion for severe dysplasia in this classification is a domed or convex trochlea. The authors have also described a treatment algorithm for PFI based on the OBC. The criteria described by the Oswestry-Bristol group have also been found to be useful by other authors. Thaunat et al[26] and Nelitz et al[27] used these to describe severity of TD in their respective case series, and Lippacher et al[28] mention these criteria to be compatible with their arthroscopic findings of TD.[26–28]

The creators of the OBC also analyzed the intra- and interobserver agreement of their classification using 32 MRIs with four observers, and found it to be better than the DJC. They found 69% agreement among the observers using the OBC, i.e. a fair to good reliability (kappa statistics), and this was even higher when expressed in terms of the nominal S-statistic.[6] The intraobserver agreement showed a significant variability using the DJC (poor to excellent) but this was much better using the OBC (fair/good to excellent). Overall, the original authors of the OBC classification report that it has a good-to-excellent intraobserver agreement and fair-to-good interobserver agreement while assessing TD. They found this to be better than the findings for the DJC on both CT and MRI.

The Oswestry Patellotrochlear Algorithm (OPTA) has been described by the authors of the OBC as a comprehensive guide to surgical management of PFI.[29] This includes the OBC and the patellotrochlear index, and has been found to be safe and effective for the same. This makes the OBC an attractive option for knee surgeons to use for classifying TD, although further clinical validation by centres around the world would help to make the OPTA more popular.

A German study by Konrads et al[30] in 2020 looked at the reliability of the OBC and found encouraging results. They observed an average interobserver agreement of > 90% (first reading 90%, second reading 96%). These were similar to our findings in the current study where we calculated the Cronbach's α values for the OBC as 0.93 and 0.945 in the two readings, respectively, suggesting excellent interobserver reliability.

The use of Cohen's kappa coefficient in this study warrants further discussion. The kappa statistic was described by Cohen[31] in 1960 to account for chance agreement between two raters, which was not possible by calculating percentage agreement alone. The problems of using kappa statistics, and the paradoxes associated with it, have been described by many studies.[32-35] One of the main issues was that, in cases with trained assessors where the raters are not merely guessing the answer, the Chance-correction offered by kappa may be too high. This would mean a low kappa value may not necessarily imply a low agreement among the raters.

Most studies similar to ours have used the kappa coefficient to assess both intraobserver and interobserver agreement. Cohen created the kappa coefficient specifically for use in binary scenarios. While it is perfectly reasonable to use it in studies with two assessors, or two separate observations for the same rater, it is not as useful in studies involving multiple raters. In this study, to avoid any confusion, we have therefore chosen to use Cronbach's α to asses inter-rater validity.

Cronbach's α was originally designed to measure and compare the internal consistency of a test or questionnaire.[21] It has since been used widely to determine interobserver reliability, especially in cases of multiple observers. Bland and Altman[36] have reported that an α score of at least 0.7 to 0.8 is required for reliability to be satisfactory. They have also stated that, in clinical applications, higher values are needed and a score of 0.9 (minimum) to 0.95 is desirable.

This study is the first comparative analysis of the intra- and interobserver reliability of the DJC and OBC by a non-founder institute. This is also, to our knowledge, one of the largest reliability studies for classification tools of TD, using 100 MRIs and assessed by six evaluators.

This study is not without its limitations. Due to the retrospective nature of this study, we have not been able to validate the clinical utility of the classification systems based on the treatment algorithms suggested, especially for the OBC. This is an area that can be targeted for future prospective research. We also feel that, though α values are better representatives of interobserver reliability, their use here has limited the comparability with other similar studies in the literature.

In conclusion, this study shows that both the DJC and OBC are reliable systems for classifying TD based on MRIs of the PFJ. The OBC has a higher intraobserver reliability and hence appears to be a more valid system. Additionally, it is simple to use and provides a robust treatment algorithm, making it an attractive classification system for knee surgeons around the world.

Take home message

- Both the Dejour and Oswestry Bristol Classification (OBC) are reliable for classifying trochlear dysplasia of the knee.

- In addition, the OBC is simple to use and provides a robust treatment algorithm, making it an attractive classification system for knee surgeons to use globally.

Twitter

Follow K. D. Roy @KRoy9010

References

1. **Batailler C**, **Neyret P**. Trochlear dysplasia: imaging and treatment options. *EFORT Open Rev*. 2018;3(5):240–247.
2. **Dejour H**, **Walch G**, **Nove-Josserand L**, **Guier C**. Factors of patellar instability: an anatomic radiographic study. *Knee Surg Sports Traumatol Arthrosc*. 1994;2(1):19–26.

3. **Lippacher S**, **Dejour D**, **Elsharkawi M**, **et al**. Observer agreement on the Dejour trochlear dysplasia classification: A comparison of true lateral radiographs and axial magnetic resonance images. *Am J Sports Med*. 2012;40(4):837–843.
4. **Nelitz M**, **Lippacher S**, **Reichel H**, **Dornacher D**. Evaluation of trochlear dysplasia using MRI: correlation between the classification system of Dejour and objective parameters of trochlear dysplasia. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(1):120–127.
5. **Kazley JM**, **Banerjee S**. Classifications in brief: The Dejour Classification of trochlear dysplasia. *Clin Orthop Relat Res*. 2019;477(10):2380–2386.
6. **Sharma N**, **Brown A**, **Bouras T**, **Kuiper JH**, **Eldridge J**, **Barnett A**. The Oswestry-Bristol Classification. *Bone Joint J*. 2020;102-B(1):102–107.
7. **Dejour D**, **Le Coultre B**. Osteotomies in patello-femoral instabilities. *Sports Med Arthrosc Rev*. 2007;15(1):39–46.
8. **Landis JR**, **Koch GG**. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
9. **van Huyssteen AL**, **Hendrix MRG**, **Barnett AJ**, **Wakeley CJ**, **Eldridge JDJ**. Cartilage-bone mismatch in the dysplastic trochlea. An MRI study. *J Bone Joint Surg Br*. 2006;88-B(5):688–691.
10. **Dejour D**, **Saggin P**. The sulcus deepening trochleoplasty-the Lyon's procedure. *Int Orthop*. 2010;34(2):311–316.
11. **Dejour H**, **Walch G**, **Neyret P**, **Adeleine P**. Dysplasia of the femoral trochlea. *Rev Chir Orthop Reparatrice Appar Mot*. 1990;76(1):45–54.
12. **Balcarek P**, **Ammon J**, **Frosch S**, **et al**. Magnetic resonance imaging characteristics of the medial patellofemoral ligament lesion in acute lateral patellar dislocations considering trochlear dysplasia, patella alta, and tibial tuberosity-trochlear groove distance. *Arthroscopy*. 2010;26(7):926–935.
13. **Colvin AC**, **West RV**. Patellar instability. *J Bone Joint Surg Am*. 2008;90-A(12):2751–2762.
14. **Diederichs G**, **Issever AS**, **Scheffler S**. MR imaging of patellar instability: injury patterns and assessment of risk factors. *Radiographics*. 2010;30(4):961–981.
15. **Koëter S**, **Bongers E**, **de Rooij J**, **van Kampen A**. Minimal rotation aberrations cause radiographic misdiagnosis of trochlear dysplasia. *Knee Surg Sports Traumatol Arthrosc*. 2006;14(8):713–717.
16. **Fucentese SF**, **von Roll A**, **Koch PP**, **Epari DR**, **Fuchs B**, **Schottle PB**. The patella morphology in trochlear dysplasia--A comparative MRI study. *Knee*. 2006;13(2):145–150.
17. **Chhabra A**, **Subhawong TK**, **Carrino JA**. A systematised MRI approach to evaluating the patellofemoral joint. *Skeletal Radiol*. 2011;40(4):375–387.
18. **Stepanovich M**, **Bomar JD**, **Pennock AT**. Are the current classifications and radiographic measurements for trochlear dysplasia appropriate in the skeletally immature patient? *Orthop J Sports Med*. 2016;4(10):2325967116669490.
19. **Remy F**, **Besson A**, **Migaud H**, **Cotten A**, **Gougeon F**, **Duquennoy A**. Reproducibility of the radiographic analysis of dysplasia of the femoral trochlea. Intra- and interobserver analysis of 68 knees. *Rev Chir Orthop Reparatrice Appar Mot*. 1998;84(8):728–733.
20. **Bujang MA**, **Omar ED**, **Baharum NA**. A review on sample size determination for Cronbach's Alpha test: A simple guide for researchers. *Malays J Med Sci*. 2018;25(6):85–99.
21. **Tavakol M**, **Dennick R**. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53–55.
22. **Carrillon Y**, **Abidi H**, **Dejour D**, **Fantino O**, **Moyen B**, **Tran-Minh VA**. Patellar instability: assessment on MR images by measuring the lateral trochlear inclination-initial experience. *Radiology*. 2000;216(2):582–585.
23. **Pfirrmann CWA**, **Zanetti M**, **Romero J**, **Hodler J**. Femoral trochlear dysplasia: MR findings. *Radiology*. 2000;216(3):858–864.
24. **Biedert RM**, **Bachmann M**. Anterior-posterior trochlear measurements of normal and dysplastic trochlea by axial magnetic resonance imaging. *Knee Surg Sports Traumatol Arthrosc*. 2009;17(10):1225–1230.
25. **Utting MR**, **Mulford JS**, **Eldridge JDJ**. A prospective evaluation of trochleoplasty for the treatment of patellofemoral dislocation and instability. *J Bone Joint Surg Br*. 2008;90-B(2):180–185.
26. **Thaunat M**, **Bessiere C**, **Pujol N**, **Boisrenoult P**, **Beaufils P**. Recession wedge trochleoplasty as an additional procedure in the surgical treatment of patellar instability with major trochlear dysplasia: early results. *Orthop Traumatol Surg Res*. 2011;97(8):833–845.
27. **Nelitz M**, **Dreyhaupt J**, **Lippacher S**. Combined trochleoplasty and medial patellofemoral ligament reconstruction for recurrent patellar dislocations in severe trochlear dysplasia: a minimum 2-year follow-up study. *Am J Sports Med*. 2013;41(5):1005–1012.
28. **Nelitz M**, **Lippacher S**. Arthroscopic evaluation of trochlear dysplasia as an aid in decision making for the treatment of patellofemoral instability. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(11):2788–2794.
29. **Sharma N**, **Rehmatullah N**, **Kuiper JH**, **Gallacher P**, **Barnett AJ**. Clinical validation of the Oswestry-Bristol Classification as part of a decision algorithm for trochlear dysplasia surgery. *Bone Joint J*. 2021;103-B(10):1586–1594.
30. **Konrads C**, **Gonser C**, **Ahmad SS**. Reliability of the Oswestry-Bristol Classification for trochlear dysplasia: expanded characteristics. *Bone Jt Open*. 2020;1(7):355–358.
31. **Cohen J**. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 1960;20(1):37–46.
32. **Feinstein AR**, **Cicchetti DV**. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*. 1990;43(6):543–549.
33. **Gross ST**. The kappa coefficient of agreement for multiple observers when the number of subjects is small. *Biometrics*. 1986;42(4):883–893.
34. **Thomsen NOB**, **Olsen LH**, **Nielsen ST**. Kappa statistics in the assessment of observer variation: the significance of multiple observers classifying ankle fractures. *J Orthop Sci*. 2002;7(2):163–166.
35. **McHugh ML**. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–282.
36. **Bland JM**, **Altman DG**. Cronbach's alpha. *BMJ*. 1997;314(7080):572.

Author information:

- K. D. Roy, MS (Orth), MRCS (Eng), Specialty Registrar in Trauma and Orthopaedics, University Hospital of Wales, Cardiff, UK.
- P. Joshi, FRCS (Tr&Orth), Specialty Doctor in Trauma and Orthopaedics
- I. Ali, MRCS (Eng), Core Surgical Trainee
- P. M. Shenoy, FRCS (Tr&Orth), Specialty Doctor in Trauma and Orthopaedics
- I. Malek, FRCS Orth, Consultant, Trauma and Orthopaedics
- D. Barlow, FRCS Orth, Consultant, Trauma and Orthopaedics
- A. Syed, FRCS Orth, Consultant, Trauma and Orthopaedics
- Y. Joshi, FRCS Orth, Consultant, Trauma and Orthopaedics
  Wrexham Maelor Hospital, Wrexham, UK.

Author contributions:

- K. D. Roy: Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing.
- P. Joshi: Conceptualization, Data curation, Project administration, Resources, Software, Validation, Writing – original draft, Writing – review & editing.
- I. Ali: Data curation, Project administration, Resources, Software, Validation, Writing – review & editing.
- P. M. Shenoy: Data curation, Project administration, Resources, Software, Validation, Writing – review & editing.
- I. Malek: Methodology, Supervision, Resources, Validation, Writing – review & editing.
- D. Barlow: Methodology, Resources, Validation, Writing – review & editing.
- A. Syed: Project administration, Validation, Writing – review & editing.
- Y. Joshi: Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Writing – original draft, Writing – review & editing.

Funding statement:

- The authors received no financial or material support for the research, authorship, and/or publication of this article.

Data sharing:

- The datasets generated and analyzed in the current study are not publicly available due to data protection regulations. Access to data is limited to the researchers who have obtained permission for data processing. Further inquiries can be made to the corresponding author.

Acknowledgements:

- The authors wish to thank Dr. Pandurang V. Thatkar (MSc, PhD), Statistician-cum-Assistant Professor at PCMC's Postgraduate Institute & YCM Hospital (Pune, India) (https://orcid.org/0000-0002-7838-8274).

Ethical review statement:

- An ethical review was not required as per HRA decision tool.

Open access funding:

- The open access fee for this study was funded by Wrexham Maelor Hospital research fund.