|  |  |
| --- | --- |
| **Title** | **The Oswestry-Bristol Classification: A New Validated assessment tool to classify trochlear dysplasia** |
| **Authors** | Sharma N1, Brown A1, Bouras T1, Kuiper J1, Eldridge J2, Barnett A1 |
| **Affiliations** | 1 Robert Jones and Agnes Hunt Orthopedic Hospital, Gobowen, UK  2 Bristol Royal Infirmary, Bristol, UK |
| **Contact** |  |

**ABSTRACT**

**Background**

Trochlear dysplasia is a significant risk factor for patellofemoral instability [1,2,3]. The Dejour classification is currently considered the gold standard for classifying trochlear dysplasia, but numerous studies reported poor results of reliability on both plain radiography and MRI [2,3,9]. The severity of trochlear dysplasia is important to establish in order to guide surgical management. We have developed an MRI specific classification system to assess the severity of trochlear dysplasia - The Oswestry-Bristol Classification. This is a 4-part classification system comprising of Normal, Mild, Moderate and Severe to represent a normal, shallow, flat and convex trochlear respectively. The purpose of this study is to assess the inter- and intraobserver reliability of the Oswestry-Bristol Classification and compare it with that of the Dejour classification.

**Patients and Methods:**

Four observers (2 senior and 2 junior orthopedic surgeons) independently assessed 32 CT and MR axial imaging for trochlear dysplasia and classified each according to the Oswestry-Bristol and the Dejour classification systems. Assessments were repeated following a 4-week interval. The inter- and intraobserver agreement was determined by using Fleiss’ generalization of Cohen’s kappa statistic and S-statistic nominal and linear weights using R vs 3.3.3 *(R Foundation for Statistical Computing, Vienna, Austria)*.

**Results**

The Oswestry-Bristol Classification showed fair to good interobserver agreement and good to excellent intraobserver agreement (mean kappa 0.68). The Dejour classification showed poor interobserver agreement and fair to good intraobserver agreement (mean kappa 0.52).

**Conclusion**

The Oswestry-Bristol Classification is a validated classification system, to assess the severity of trochlear dysplasia. It can be utilized in clinical practice to simplify and standardize surgical decision-making in patients with recurrent patella instability.

**MANUSCRIPT BODY**

**Introduction:**

Trochlear dysplasia is a significant risk factor in recurrent patellofemoral instability [1,2,3] and is found in up to 96% of patients’ [4]. It was first described by Maldague and Malghem in 1985 [5]. H. Dejour introduced a 3-grade classification system in 1990 [1] and was subsequently modified by D. Dejour in 1998 to include 4-grades in an effort to improve the intra- and inter observer agreement [6]. This classification includes four morphological grades of increasing severity (A to D) on both 2-dimensional radiographs and computerized tomography (CT). Dejour proposed a treatment algorithm based on the grade of trochlear dysplasia.

Assessing trochlear dysplasia on a lateral radiograph is challenging due to the difficulties in obtaining a true lateral image, and whilst CT scans are excellent at demonstrating bony anatomy, they do not show cartilage structure, which contributes to the shape of the trochlea [7]. Magnetic resonance (MR) imaging has superseded plain radiographs and CT, and is now considered the gold standard to assess the severity of trochlear dysplasia [6]. MR imaging provides information on the integrity of the articular cartilage and medial patellofemoral ligament (MPFL), any associated osteochondral injury, patella height measurements and tibial tuberosity-trochlear groove distance (TTTG) [8].

The Dejour classification has been used on MRI, but poor inter- and intraobserver agreement has been reported [2,3,9] with low correlation to objective intraoperative findings [6]. A more simplified approach of grading trochlear dysplasia has been called for [10].

We have developed a new classification system to assess the severity of trochlear dysplasia – the Oswestry-Bristol Classification (OBC). This classifies dysplasia into four grades based on axial MR imaging: Normal, Mild, Moderate and Severe which represent a normal, shallow, flat and convex trochlear respectively (Figure 1). The purpose of this study was to determine inter- and intraobserver agreement of the Oswestry-Bristol Classification and compare it with the Dejour classification system.

**Patients and Methods:**

This study was conducted at the Robert Jones and Agnes Hunt Orthopedic Hospital, Oswestry, UK. A retrospective review of a prospectively maintained longitudinal patellofemoral database was performed. Patients who had undergone both CT and MRI scanning to investigate recurrent patellofemoral instability were identified. We evaluated 32 CT and MRI scans in 28 patients, none of which had undergone previous surgery.

Each observer was provided with a detailed illustrated description of both the Dejour and OBC classification systems. Each image was read twice, four-weeks apart by four different Orthopedic Surgeons (2 senior and 2 junior). An interlude of at least 1 hour was made between grading CT and MR imaging. The Dejour classification has been well described in the literature [6] and was used to grade both CT and MR images.

The Oswestry-Bristol Classification

Magnetic Resonance Imaging of the knee is performed under standard conditions, using a Simemens Skyra 3 Tesla Magnetom. As this classification assesses the bony-cartilaginous architecture, the classification can be made irrespective of knee position (e.g. extended versus 30 degrees flexed, supine vs standing).

The axial imaging of the fat-saturated sequences is used for assessment of the morphological appearance of the trochlea. Scrolling from superior to inferior, the first cranio-caudal image where the trochlea cartilage can be reliably visualized is selected; this image is then used to classify the severity of trochlea dysplasia (Figure 1).

**Statistical analysis**

The inter- and intraobserver agreement was assessed using Fleiss’ generalization of Cohen’s kappa statistic [11]. The kappa statistic expresses the chance-corrected agreement; it is the (normalized) observed agreement minus the agreement expected on the basis of chance alone. The expected agreement is based on the prevalence of each grade, which was calculated from the combined ratings of all raters. Values of kappa greater than 0.75 or so may be taken to represent excellent agreement beyond chance, values below 0.40 or so may

be taken to represent poor agreement beyond chance, and values between 0.40 and 0.75 may be taken to represent fair to good agreement beyond chance [12]. For the intra-rater agreement, the mean kappa of all observers was determined [13].

The exact value of kappa is influenced not only by the agreement but also by the relative prevalence of the grades, alternative formulations of kappa have been proposed. One such measure, the S-statistic, was introduced recently [14]. It seems to address the problems with kappa, and was therefore also calculated for this dataset. The S-statistic can be calculated in two forms, a nominal and ordinal one. The nominal form, like Fleiss’ kappa statistic, does not take into account the ordered nature of the dysplasia grades, whereas the ordinal form weighs disagreement by the number of grades that the raters depart from each other. We used a linear weighting scale in our calculations. Confidence intervals and p-values for the difference between the 4-type Dejour and the Oswestry-Bristol Classification grading systems were calculated using a normal approximation from 999 ordinary bootstrap samples. All statistical analyses were done using R vs 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) using the packages *irr*, *raters* and *boot*.

**Results**

CT and MR imaging of 32 knees in 28 patients (16 female, mean age of 24y, SD 10y, range 12-48y) were assessed. Based on the ratings of all four surgeons using the OBC, 5% were classified as “Normal”, 23% “Mild”, 52% “Moderate”, and 21% “Severe”. Using the MR based Dejour classification, 5% were classified as “Normal”, 20% grade A, 42% grade B, 16% grade C and 17% grade D.

*Interobserver agreement*Agreement, kappa-and S-statistic values were calculated separately for each set of observations (Tables 1 and 2). The values of all statistics were consistently lower in the second round of observations. In the first round, the observed agreement for MRI was 69% for the Oswestry-Bristol Classification compared with 55% for the Dejour classification (Table 1).

When expressed using the kappa statistics, the reliability was fair to good for the OBC compared with poor for the Dejour classification (Table 1). Expressed in terms of the nominal S-statistic, the reliability was slightly higher. However, when accounting for the ordinal nature of the grading systems by using a linear weighting, the S-statistic suggested that both grading systems had a fair to good reliability in both rounds (Table 1 and 2). When measured using the S-statistic, the reliability of the OBC grading was larger than that of the Dejour grading (Table 1)

*Intraobserver agreement*The mean intraobserver agreement was fair to good for both systems (Table 3). There was a wide range of individual observers’ agreements, ranging from poor to excellent (Table 3). When using the Oswestry-Bristol Classification system, all raters achieved fair to good or even excellent agreement.

**Discussion**

The principal finding of this study is that the Oswestry-Bristol Classification demonstrates fair to good inter-observer agreement and good to excellent intra-observer agreement in the assessment of trochlea dysplasia. These were both better than that of the Dejour classification on both CT and MRI, which demonstrated poor inter-observer agreement and fair to good intra-observer agreement

The Oswestry-Bristol Classification has four grades: Normal, Mild, Moderate and Severe based on the morphological appearance of the femoral trochlear on axial MR scans. Classifying trochlear dysplasia simplifies the surgical decision-making process of a complex problem and aids future research. We have developed a surgical algorithm for managing recurrent patellofemoral instability based on the OBC along with patella height (Figure 2).

The poor inter- intraobserver agreement found in this study with the Dejour classification is corroborated in several recent studies within the literature. Remy et al [3] studied 68 lateral radiographs using 7 observers and typed the trochlear dysplasia according to the 4-type Dejour classification. In their study none of the 68 lateral radiographs were recognized as having the same shape by the 7 observers and concluded a low interobserver agreement.

Lippacher et al [9] evaluated the intra- and interobserver agreements of using the 4-type Dejour classification for both plain radiography (24-78%) and MR imaging (25-81%). The authors found the Dejour classification was insufficient to grade the severity of trochlear dysplasia on MR imaging. Stepanovich et al [10] analyzed 36 lateral radiographs and axial MR scans of skeletally immature patients’ with patellofemoral instability and graded the degree of trochlear dysplasia using the Dejour classification, trochlear depth index, lateral trochlear inclination and medial condyle trochlear offset. They found only 42% of patients’ had an adequate lateral radiograph, where the Dejour classification could be reliably used. In addition the Dejour classification on MR imaging had the poorest intra and interobserver agreement when compared to all other radiographic parameters. They concluded that Dejour’s classification had “suboptimal reliability” on MR imaging. In our study, we found the Dejour classification had poor inter-observer agreement on MR imaging and the same intra-observer agreement for both CT and MR imaging.

Tscholl et al [15] showed the Dejour classification has only fair agreement between using plain radiography and MR imaging and concluded the classification tended to underestimate the severity of dysplasia.

Other methods to quantify the degree of trochlear dysplasia are well described and include the femoral sulcus angle lateral trochlear inclination angle, trochlea depth index and medial condyle trochlear offset.

Carillon et al [16] found the lateral trochlear inclination (LTI) of less than 11°on MR imaging has a sensitivity, specificity and accuracy of 93%, 87% and 90% respectively, in diagnosing patellofemoral instability. Pfirrmann et al [17] used objective parameters and demonstrated that trochlear dysplasia could be reliably diagnosed using MR images greater than 3cm above the joint line. Biedert and Bachmann [18] measured the trochlear, medial and lateral condylar height and compared this to the total width of the femoral condyle, on axial MR imaging. They found that in 5 out of 6 cases the pathology was located in the centre and/or medial trochlear.

Nelitz et al [2] studied the correlation between Dejour and objective parameters as stated by Carillon et al [16], Pfirrmann et al [17] and Biedert and Bachmann [18]. 80 MR images were analyzed. They found none of the objective parameters could be correlated with Dejour.

Zimmerer et al [19] reported the need for a reliable and valid classification to grade trochlear dysplasia and recommended that the Dejour classification be only grouped into low grade (Type A) and high grade (Type B-D).

There are several limitations to this study. The main purpose of the study was to assess the reliability of the OBC. The concurrent assessment of the Dejour classification may have introduced assessor bias as a competing classification system. Nevertheless, the study has validated the OBC regardless of our results of the Dejour classification reliability. Our results of the intra- and interobserver reliability of the Dejour classification compare to that in the published literature. Secondly, we found only fair to good interobserver reliability in the OBC. However, this was between a total of four surgeons, including two Consultant Surgeons and two non-specialist registrars. There was a trend to improved inter- and intra-observer agreement between the two senior surgeons compared to the junior surgeons. We anticipate the OBC will be used by experts in the surgical decision-making process, and therefore expect that the inter-observer reliability may well be better than found in this study.

**Conclusion**

The Oswestry-Bristol Classification is a validated grading system to categorize the severity of trochlear dysplasia. It is simple and reproducible and has better inter- and intraobserver agreement than that of the Dejour classification.

**Take home message:**

The Oswestry-Bristol Classification is a simple, reproducible and valid assessment tool to classify trochlear dysplasia. It provides a straightforward treatment algorithm for the surgical management of recurrent patellofemoral instability.

**References:**

1] Dejour H, Walch G. Neyret PH, Adeleine P. La dysplasia de la trochlee femorale. Rev Chir Orthop Reparatrice Appar Mot. 1990;76:45-54

2] 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 surgery, sports traumatology, arthroscopy Jan 2014; vol. 22 (no. 1); p. 120-127

3] Rémy, F; Chantelot, C; Fontaine, C; Demondion, X; Migaud, H; Gougeon, F. Inter- and intraobserver reproducibility in radiographic diagnosis and classification of femoral trochlear dysplasia. Surgical and radiologic anatomy : SRA; 1998; vol. 20 (no. 4); p. 285-289

4] Dejour H, Walch G et al. Factors of patella instability: an anatomic radiographic study. Knee Surg Sports Traumatol Arthrosc. 1994;2:19-26

5] Maldague B, Malghem J. Apport du cliché de profil du genou dans le depistage des instabilities rotuliennes: rapport preliminaire. Rev chir orthop. 1985; 71:5-13

6] Nelitz, M; Lippacher, S. Arthroscopic evaluation of trochlear dysplasia as an aid in decision making for the treatment of patellofemoral instability. Knee surgery, sports traumatology, arthroscopy. Nov 2014; vol. 22 (no. 11); p. 2788-2794

7] [Koëter S](https://www.ncbi.nlm.nih.gov/pubmed/?term=Ko%C3%ABter%20S%5BAuthor%5D&cauthor=true&cauthor_uid=16395562)1, [Bongers EM](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bongers%20EM%5BAuthor%5D&cauthor=true&cauthor_uid=16395562), [de Rooij J](https://www.ncbi.nlm.nih.gov/pubmed/?term=de%20Rooij%20J%5BAuthor%5D&cauthor=true&cauthor_uid=16395562), [van Kampen A](https://www.ncbi.nlm.nih.gov/pubmed/?term=van%20Kampen%20A%5BAuthor%5D&cauthor=true&cauthor_uid=16395562). Minimal rotation aberrations cause radiographic misdiagnosis of trochlear dysplasia. [Knee Surg Sports Traumatol Arthrosc.](https://www.ncbi.nlm.nih.gov/pubmed/16395562) 2006 Aug;14(8):713-7. Epub 2006 Jan 5.

8] Shen J.; Qin L.; Yao W.-W.; Li M. The significance of magnetic resonance imaging in severe femoral trochlear dysplasia assessment. Experimental and Therapeutic Medicine; Dec 2017; vol. 14 (no. 6); p. 5438-5444

9] Lippacher, Sabine; Dejour, David; Elsharkawi, Mohammed; Dornacher, Daniel; Ring, Christina; Dreyhaupt, Jens; Reichel, Heiko; Nelitz, Manfred. Observer agreement on the Dejour trochlear dysplasia classification: a comparison of true lateral radiographs and axial magnetic resonance images.

10] Stepanovich, Matthew; Bomar, James D; Pennock, Andrew T. Are the Current Classifications and Radiographic Measurements for Trochlear Dysplasia Appropriate in the Skeletally Immature Patient? Orthopaedic journal of sports medicine; Oct 2016; vol. 4 (no. 10); p. 2325967116669490

11] Shoukri MM. Measures of interobserver agreement and reliability. CRC press; 2004.

12] Fleiss JL, Levin B, Paik MC. Statistical methods for rates and proportions; 3rd Edition. John Wiley & Sons; 2003.

13] Light RJ. Measures of response agreement for qualitative data: some generalizations and alternatives. Psychological bulletin. 1971;76(5):365.

14] Marasini D, Quatto P, Ripamonti E. Assessing the inter-rater agreement for ordinal data through weighted indexes. Statistical methods in medical research. 2016;25(6):2611-33.

15] Tscholl. Conventional radiographs and magnetic resonance imaging for the analysis of TD.

16] [Carrillon Y](https://www.ncbi.nlm.nih.gov/pubmed/?term=Carrillon%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=10924589)1, [Abidi H](https://www.ncbi.nlm.nih.gov/pubmed/?term=Abidi%20H%5BAuthor%5D&cauthor=true&cauthor_uid=10924589), [Dejour D](https://www.ncbi.nlm.nih.gov/pubmed/?term=Dejour%20D%5BAuthor%5D&cauthor=true&cauthor_uid=10924589), [Fantino O](https://www.ncbi.nlm.nih.gov/pubmed/?term=Fantino%20O%5BAuthor%5D&cauthor=true&cauthor_uid=10924589), [Moyen B](https://www.ncbi.nlm.nih.gov/pubmed/?term=Moyen%20B%5BAuthor%5D&cauthor=true&cauthor_uid=10924589), [Tran-Minh VA](https://www.ncbi.nlm.nih.gov/pubmed/?term=Tran-Minh%20VA%5BAuthor%5D&cauthor=true&cauthor_uid=10924589). Patellar instability: assessment on MR images by measuring the lateral trochlear inclination-initial experience.[Radiology.](https://www.ncbi.nlm.nih.gov/pubmed/10924589) 2000 Aug;216(2):582-5.

# 17] [Pfirrmann CW](https://www.ncbi.nlm.nih.gov/pubmed/?term=Pfirrmann%20CW%5BAuthor%5D&cauthor=true&cauthor_uid=10966723)1, [Zanetti M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Zanetti%20M%5BAuthor%5D&cauthor=true&cauthor_uid=10966723), [Romero J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Romero%20J%5BAuthor%5D&cauthor=true&cauthor_uid=10966723), [Hodler J](https://www.ncbi.nlm.nih.gov/pubmed/?term=Hodler%20J%5BAuthor%5D&cauthor=true&cauthor_uid=10966723). Femoral trochlear dysplasia: MR findings. [Radiology.](https://www.ncbi.nlm.nih.gov/pubmed/10966723) 2000 Sep;216(3):858-64.

18] [Biedert RM](https://www.ncbi.nlm.nih.gov/pubmed/?term=Biedert%20RM%5BAuthor%5D&cauthor=true&cauthor_uid=19495725)1, [Bachmann M](https://www.ncbi.nlm.nih.gov/pubmed/?term=Bachmann%20M%5BAuthor%5D&cauthor=true&cauthor_uid=19495725). Anterior-posterior trochlear measurements of normal and dysplastic trochlea by axial magnetic resonance imaging.[Knee Surg Sports Traumatol Arthrosc.](https://www.ncbi.nlm.nih.gov/pubmed/19495725) 2009 Oct;17(10):1225-30. doi: 10.1007/s00167-009-0824-y. Epub 2009 Jun 4.

19] [Alexander Zimmerer](https://www.ncbi.nlm.nih.gov/pubmed/?term=Zimmerer%20A%5BAuthor%5D&cauthor=true&cauthor_uid=29322270), [Christian Sobau](https://www.ncbi.nlm.nih.gov/pubmed/?term=Sobau%20C%5BAuthor%5D&cauthor=true&cauthor_uid=29322270), and [Peter Balcarek](https://www.ncbi.nlm.nih.gov/pubmed/?term=Balcarek%20P%5BAuthor%5D&cauthor=true&cauthor_uid=29322270). Recent developments in evaluation and treatment of lateral patellar instability. J Exp Orthop. 2018 Dec; 5: 3.

Figure 1: The Oswestry-Bristol Classification. Axial, fat saturated images obtained at the level of the first craniocaudal image where the trochlea cartilage can be reliably visualized. A. Normal trochlea. B. Mild trochlea dysplasia – shallow trochlea groove. C. Moderate dysplasia – flat trochlea. D. Severe dysplasia – convex trochlea.



Figure 2. Algorithm for the surgical management of recurrent patellofemoral instability based on the OBC. Indications for additional TTD: patellotrochlear index less than 20%

MPFL: medial patellofemoral ligament, TTD: tibial tubercle distalization.



Table 1 Interobserver agreement, Kappa and S statistic values for first round of observations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 4-type Dejour grade | | OBC grade | Difference OBC-Dejour | |
|  | CT | MRI | MRI | Mean (95% CI) | p-value |
| Observed agreement | 49% | 55% | 69% |  |  |
| Expected agreement | 26% | 27% | 36% |  |  |
| Kappa  (95% CI) | 0.32  (0.24-0.40) | 0.38  (0.30-0.46) | 0.51  (0.41-0.60) | 0.13  (-0.02-0.14) | 0.09 |
| S-statistic (nominal) (95% CI) | 0.37  (0.24-0.52) | 0.44  (0.34-0.54) | 0.58  (0.46-0.72) | 0.14  (0.03-0.26) | 0.01 |
| S-statistic (linear weights) (95% CI) | 0.57  (0.45-0.70) | 0.67  (0.58-0.74) | 0.74  (0.66-0.83) | 0.07  (0.01-0.14) | 0.03 |

Table 2 Interobserver agreement, Kappa and S statistic values for second round of observations

|  |  |  |  |
| --- | --- | --- | --- |
|  | 4-type Dejour grade | | OBC |
|  | CT | MRI | MRI |
| Observed agreement (%) | 38 | 45 | 58 |
| Expected agreement (%) | 27 | 29 | 37 |
| Kappa (95% CI) | 0.16 (0.07-0.24) | 0.22 (0.14-0.31) | 0.34 (0.24-0.44) |
| S-statistic (nominal) | 0.23 (0.11-0.35) | 0.31 (0.21-0.43) | 0.44 (0.31-0.57) |
| S-statistic (linear weights) | 0.46 (0.35-0.58) | 0.56 (0.47-0.65) | 0.66 (0.57-0.75) |

Table 3 Mean and range of intraobserver Kappa values

|  |  |  |
| --- | --- | --- |
|  | Mean kappa | Range (Rater) |
| Dejour (CT) | 0.52 | 0.36 (R4) to 0.70 (R3) |
| Dejour (MRI) | 0.52 | 0.27 (R1) to 0.78 (R3) |
| OBC (MRI) | 0.68 | 0.41 (R1) to 0.80 (R2) |