*Acta Orthop. Belg.*, 2022, **88**, 805-810

*ORIGINAL STUDY*

*doi.org/ 10.52628/88.4.9845*

# An evaluation of intra and inter observer reliability of the five used classification systems of tibial plateau fractures

Alper Koksal, Kadir Abul, Deniz Kargin, Osman Cimen, Muhammed Mert, Abdi Keskin

*From the Department of Orthopedics and Traumatology, Baltalimani Bone Diseases Training and Research Hospital, Istanbul, Turkey*

**We aimed to evaluate the intraobserver and interobserver variations of the five primary classification systems for assessing tibial plateau fractures via standard X-Ray, biplanar and reconstructed 3D CT images. Using anteroposterior (AP) – lateral X-Ray, and CT images, one hundred tibial plateau fractures were evaluated and classified by four surgeons according to the AO, Moore, Schatzker, modified Duparc, and 3-column classification systems. Each observer evaluated the radiographs and CT images separately – listed each time randomly – on a total of 3 occasions: with an initial evaluation, and then subsequently in weeks 4 and 8. Intra- and interobserver variabilities were assessed using the Kappa statistics. Intra- and interobserver variabilities were 0.55 ± 0.03 and 0.50 ± 0.05 for AO, 0.58 ± 0.08 and 0.56 ± 0.02 for Schatzker, 0.52 ± 0.06 and 0.49 ± 0.04 for Moore, 0.58 ± 0.06 and 0.51± 0.06 for the modified Duparc, and 0.66 ± 0.03 and 0.68 ± 0.02 for the 3-column classification. Evaluation of tibial plateau fractures using 3-column classification in conjunction with radiographic classifications has higher levels of consistency compared to radiographic classifications alone.**

**Keywords:** Tibial plateau fractures; 3-column; interobserver; intraobserver; reliability.

## INTRODUCTION

To determine a suitable treatment method for a fracture, it is necessary to define the pathology properly. Fractures of the tibial plateau must be defined in detail before they can be treated. The difficulty in identifying tibial plateau fractures radiographically led to the establishment of many classification schemes. Bradfort et al. made a simple classification in 1950 *(1)*. Hohl and Luck conducted a study in 1956 to classify tibial plateau fractures *(2)*. Rasmussen sorted tibial plateau fractures into three main groups (lateral condyle, medial condyle, and double condyle fractures) in 1973 *(3)*. The most

■ Alper Koksal[1],
■ Kadir Abul[2],
■ Deniz Kargın[1],
■ Osman Cimen[1],
■ Muhammed Mert[1],
■ Abdi Keskin[3].

*[1]Department of Orthopedics and Traumatology, Baltalimani Bone Diseases Training and Research Hospital, Baltalimanı Hisar Caddesi 62, 34470, Istanbul, Turkey.*
*[2]Department of Orthopedics and Traumatology, Basaksehir Pine and Sakura State hospital, Başakşehir Mahallesi G-434 Caddesi No:2L 34480, Başakşehir/Istanbul Turkey.*
*[3]Department of Orthopedics and Traumatology, Nizip State Hospital, Fatih Sultan Mahallesi, Hastane Caddesi,27400, Gaziantep, Turkey.*

Correspondence : Alper Koksal, Department of Orthopedics and Traumatology, Baltalimani Bone Diseases Training and Research Hospital, Baltalimanı Hisar Caddesi 62, 34470, Istanbul, Turkey. Phone: +90 506 8547255. Fax: +90 212 3237090.
Email : dralperkoksal@gmail.com


*No financial support was received in the conduct of this study. There are no conflicts of interest.*

broadly accepted classification for these fractures is the well-known Schatzker classification system *(4)*. Moore published his own "Fracture-Dislocation" Classification in 1981 *(5)*. The Duparc classification was used in France *(6)*. Later, Gicquel et al. Revised the Duparc classification *(6)*. The AO/OTA (Arbeitsgemeinschaft für Osteosynthesefragen / Orthopaedic Trauma Association) classification is frequently utilized to identify tibial plateau fractures *(7)*. The widespread use of computerized tomography (CT) is beneficial for the detailed evaluation of the fracture fragments and treatment planning *(7)*. Luo et al. evaluated a multi-planar CT image of the tibial plateau fractures and reported a newer classification in 2010 whereby they divided the anatomical plateau into three main columns *(8)* (Figure 1). This method has gained increasing popularity in recent years. The numerous classification systems defined by far indicate that the definition of the tibia plateau fractures requires a common language. The joint line fractures require special attention in both definition and surgical planning *(6-8)*. Although several studies report the observer reliability of those classifications, there are insufficient data in the literature to assess the intra- and interobserver variability of all classification modalities for tibial plateau fractures, and hence, this study was conducted. We sought to evaluate the intra- and interobserver variability of the five used classification systems for the plateau fractures (Schatzker, AO/OTA, Moore, revised Duparc, and 3-column classifications) by plain radiographs and CT images.

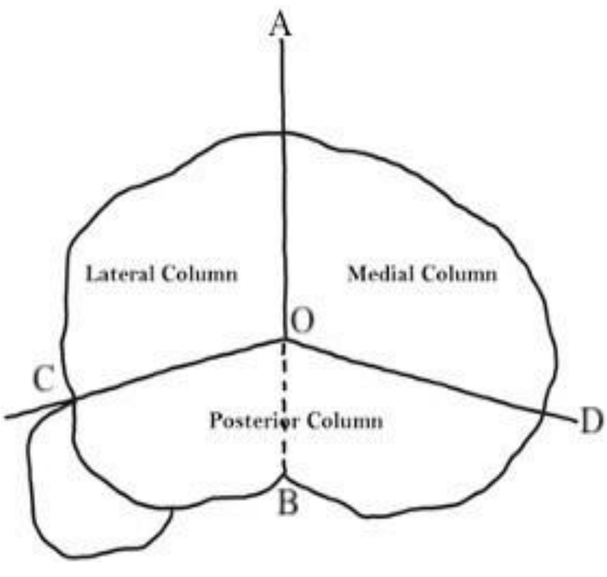


***Figure 1.*** — 3-column classification based on the transverse CT view, the tibial plateau is divided into three areas: lateral, medial, and posterior column. These three columns are separated by three connecting lines, namely OA, OC, and OD. Accordingly, Point O is the midpoint of the tibial eminences; Point A refers to anterior tibial tuberosity; Point D is the proximal tibia posteromedial ridge; Point B depicts posterior tibial sulcus, and Point C refers to the most anterior part of the head of the fibula.

## MATERIAL AND METHODS

This retrospective study was performed after obtaining the approval of the Ethical Review Board and conducted under the *Declaration of Helsinki*. Patients who underwent surgery (open reduction and internal fixation) for fractures of the tibial plateau at a tertiary referral hospital from 2012-2017 were identified through our medical records review**.** Patients with pathological fractures (n = 10), patients with low-quality radiographic images (n=15), and patients whose preoperative CT scans were not available (n=17) were excluded. A total of 100 patients were involved in the current study.

One folder **(R)** containing only preoperative AP and lateral radiographs, another folder **(CT)** containing just the preoperative CT images, and a third folder **(R / CT)** containing preoperative AP and lateral X-Ray images and CT images for each case were prepared. Patient information in the three folders was accessed only by the researchers in this study and numbered randomly. Four observers, who themselves never met with these patients (2 intermediate [2-4 year] senior training orthopedic surgeons, one advanced senior [5 years] training orthopedic surgeon, and one orthopedic surgeon experienced in trauma) working in the same clinic were assigned to evaluate radiographs and CTs. The observers had attended a brief training course on the classification methodology before the evaluation. The participating clinicians were requested to assess the images according to the Schatzker, AO / OTA, revised Duparc, Hohl and Moore, and 3-column classifications. The Schatzker classification utilized six types, the AO/OTA classification utilized six types (excepting extra-articular fractures), the Hohl and Moore classification had five types, and the Duparc classification consisted of 4 fracture types and 16 sub-types. Sample classification charts were given to the observers to ensure standardization in the evaluation of radiographs and CTs. The observers were asked to evaluate the CT images according to the 3-column classification defined by Luo et al. The CT-based classification is based on the transverse view; the tibial plateau is separated into three areas: medial, lateral, and the posterior column. The reference points demonstrating the borders of these columns are defined as **point O** (midpoints of two tibial spines), **point A** (anterior tibial tuberosity), **point D** (the posteromedial ridge of the proximal tibia), and **point C** (the most anterior point of the fibular head). On the other hand, **point B** is depicted as the posterior sulcus of the tibial plateau that divides the posterior column as the medial and lateral sections *(8)*. A separate chart and numbering, independently from the radiographic images were created for the evaluation of CT images. Each observer evaluated the radiographs and CT images – listed each time differently – on a total of 3 occasions: during an initial evaluation and then subsequently in weeks 4 and 8. After the first evaluations by the observers, no feedback was provided, and the radiographic images were not made available to any of the participants between the first, second, and third evaluation periods.

In evaluating the reliability of a fracture classification system accurately, Cohen's kappa coefficient statistic considers the level of agreement one would expect based on chance alone. The value of kappa ranges from -1 (showing complete disagreement) to 1 (showing complete agreement). A kappa value of 0 represents an agreement purely

Table I.

**a. κ value for tibia plateau fracture classifications discussed in the study**

| | **AO Classification** | **Schatzker Classification** | **Moore Classification** | **Duparc Classification** | **3-Column Classification** |
|---|---|---|---|---|---|
| **The mean interobserver κ value for all observers** | 0.50 | 0.56 | 0.49 | 0.51 | **0.68** |
| **The mean intraobserver κ value for all observers** | 0.55 | 0.58 | 0.52 | 0.58 | **0.66** |

**b. Comparison of Schatzker and 3-Column Classifications**

| | **Interobserver** | | | | | | | | **Intraobserver** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **A** | | **B** | | **C** | | **D** | | | |
| | **Schatzker** | **3-Column** | **Schatzker** | **3-Column** | **Schatzker** | **3-Column** | **Schatzker** | **3-Column** | **Schatzker** | **3-Column** |
| **A** | | | 0.57 | 0.66 | 0.54 | 0.68 | 0.55 | 0.69 | 0.51 | 0.62 |
| **B** | | | | | 0.58 | 0.67 | 0.58 | 0.71 | 0.63 | 0.68 |
| **C** | | | | | | | 0.54 | 0.67 | 0.66 | 0.64 |
| **D** | | | | | | | | | 0.51 | 0.68 |
| **The Mean κ value for each observer** | 0.51 | 0.62 | 0.54 | 0.63 | 0.55 | 0.65 | 0.56 | 0.69 | | |

by chance. The Landis and Koch criteria were used to interpret these kappa values *(9)*.

GraphPad Prism v8.0 (GraphPad Software for Windows, La Jolla California USA) was used for the statistical analyses. Intraobserver reliability was determined using Cohen's kappa value and was calculated for each observer separately *(10)*.

Fleiss' kappa was used for the interobserver reliability calculation using a multi-rater combined kappa value test *(11)*. The Fleiss' kappa value was calculated separately for each scoring session and each classification system.

Fischer's exact t-test was used for the compatibility analysis of different classification methods on fracture evaluation. A p-value lower than 0.05 was considered statistically significant for this study.

## RESULTS

The comparative evaluation of κ statistical analysis of interobserver and intraobserver variations was summarized (Table I a). Table I b demonstrated the κ statistical analysis results of intra- and interobserver variations for the most commonly used Schatzker and CT based 3-column classifications (Table I b).

Fort the AO classification, the overall mean κ value for the inter-observer variation was 0.50 ± 0.05 (range:0.43-0.56), and the intraobserver analysis gave κ values a mean of 0.55 ± 0.03 (range:0.52-0.60).

For the Schatzker classification, the overall mean κ value for the inter-observer variation was 0.56 ± 0.02 (range:0.54-0.58), and the intraobserver analysis gave κ values a mean of 0.58 ± 0.08 (range: 0.51-0.66).

For the Moore classification, the overall mean κ value for the inter-observer variation was 0.49 ± 0.04 (range:0.43-0.53), and the intraobserver analysis gave κ values a mean of 0.52 ± 0.06 (range:0.46-0.59).

For the Duparc classification, the overall mean κ value for the inter-observer variation was 0.51 ± 0.06 (range: 0.41-0.57), and the intraobserver analysis gave κ values a mean of 0.58 ± 0.06 (range:0.53-0.67).

For the 3-column classification, the overall mean κ value for the inter-observer variation was 0.68 ± 0.02 (range:0.66-0.71), and the intra-observer analysis gave κ values a mean of 0.66 ± 0.03 (range:0.62-0.68). A statistically significant difference was found in the intra- and interobserver variation between the 3-column classification and the other classification systems ($p < 0.05$).

Regarding Landis and Koch criteria, 3-column classification showed substantial agreement and, intra- and interobserver reliability was higher in 3-column than the others.

## DISCUSSION

The classification of radiographic evaluations with the AO/OTA, Schatzker, Hohl and Moore, and revised Duparc classification systems may result in variable intra- and interobserver results. Our study demonstrated that 3-column classification showed a substantial intra- and inter-observer agreement compared to radiographic evaluations demonstrating moderate agreement rates. Even though these five classification schemes have been utilized for tibial plateau fractures, neither one has been globally accepted.

Gicquel et al. found that interobserver reproducibility for X-ray image analysis lacked for the Duparc (Duparc κ = 0.365). They also stated that the reproducibility improved in-depth when CT scans were also included (Duparc κ = 0.474) *(6)*. Taşkesen et al. also found moderate interobserver agreement in their study using Duparc classification *(12)*. The κ statistics for the updated Duparc classification in the current study were similar to those seen in the literature for interobserver reliability. The likely rationale for this same moderate agreement may be that we also used CT scans along with X-ray images.

Kappa statistics are most broadly obtainable in the printed data for the Schatzker classification system. While our results were similar to quoted κ results from some studies, we detected worse findings than Brunner and Hu et al.; and better findings than Mellama et al. *(13-17)*. The κ statistics obtained for the AO/OTA classification in the current study were similar (κ 0,50) for interobserver reliability to previously reported studies *(6,12)*.

Zhu Y et al. retrospectively compared tibial plateau fractures with Schatzker, AO/OTA, and 3-column classifications and reported that the interobserver reliability for 3-column classification is superior to that of other classifications. The mean kappa value for interobserver reliability concerning the 3-column classification system was 0.766 (range: 0.706-0.890), representing substantial agreement *(18)*. However, they assessed a low number of cases (n = 50) and compared only three classifications. Mellema et al. reported fair interobserver agreement for the Schatzker and 3-column classification in their study. They found that even the addition of 3D CT reconstruction did not change the overall interobserver reliability *(17)*. However, they also investigated their study on a considerably lower number of cases (n = 15) and analyzed only these two classifications, and only the interobserver reliability. Taşkesen et al. also found moderate interobserver agreement in their study using the Schatzker, AO, 3-column, Moore, and Duparc classifications *(12)*. In the present study, interobserver reliability was substantial for 3-column and moderate for the Duparc, Schatzker, AO/OTA, Hohl and Moore classifications. There are also studies in the literature assessing intraobserver reliability. In one of these studies, Billi Am et al. found an almost perfect agreement for Schatzker, AO/OTA, and 3-column, along with moderate agreement for the Duparc classification *(19)*. Gicquel et al. found substantial intraobserver correlations for Schatzker, AO/OTA, and Duparc classifications *(6)*.

Millar SC reported that the reliability of more simplistic methods like 3-column classification was typically high (intraobserver -$\kappa$ = 0.67-0.81, interobserver -$\kappa$ = 0.71-0.87) in their literature review study *(20)*. In the current investigation, intraobserver reliability was substantial for 3-column and moderate for other classifications.

Clinical experience has been proposed as a factor affecting the interobserver reliability between participating clinicians with varying levels of expertise *(15,21)*. Although the observers also had different levels of clinical experience in our study, we found no significant difference among the participants. Evaluation of tibial plateau fractures using 3-column classification in conjunction with radiographic classifications has higher levels of consistency compared to radiographic classifications alone, regardless of experience level.

There are some potential limitations of this study that require careful consideration. First, the number of cases included seems low-though still higher than in previously reported studies. Second, images were evaluated by four observers. More valid results may be obtained with a higher number of observers. The main strength of the present study is that our study includes 100 patients, which evaluated alongside all radiographic classification methods comparatively. When evaluating cases, not only with 3D-CT images, we also assessed them together with the X-Ray images.

## CONCLUSION

The 3-column classification system for tibial plateau fractures gives a higher understanding of fracture type and an initial radiographic classification. The classification of radiographic evaluations with given systems may result in variable inter- and intraobserver results. Evaluation of 3-column classification of tibial plateau fractures together with radiographic classifications has higher levels of consistency compared to radiographic classifications alone, regardless of experience level.

## REFERENCES

1. **Bradford CH, Kilfoyle RM.** Fractures of the lateral tibial condyle. J Bone Joint Surg Am. 1950;32:39-47.
2. **Hohl M, Luck JV.** Fractures of the tibial condyle; a clinical and experimental study. J Bone Joint Surg Am. 1956;38:1001-18.
3. **Rasmussen PS.** Tibial condylar fractures. Impairment of knee joint stability as an indication for surgical treatment. J Bone Joint Surg Am. 1973;55:1331-50.
4. **Schatzker J, McBroom R, Bruce D.** The tibial plateau fracture. The Toronto experience 1968--1975. Clin Orthop Relat Res. 1979;138:94-104.
5. **Moore TM.** Fracture--dislocation of the knee. Clin Orthop Relat Res. 1981;156:128-40.
6. **Gicquel, T., Najihi, N., Vendeuvre T, Teyssedou S, Gayet L-E, Huten D.** Tibial plateau fractures: reproducibility of three classifications (Schatzker, AO, Duparc) and a revised Duparc classification. Orthop Traumatol Surg Res. 2013;99:805-16.

7. **Muller ME, Allgower M, Schneider R, Willenegger H.** Manual of Internal Fixation (3rd ed.) Heidelberg, Springer, 1991,142-143.
8. **Luo, C. F., Sun, H., Zhang, B., Zeng B-F.** Three-column fixation for complex tibial plateau fractures. J Orthop Trauma. 2010;24:683-92.
9. **Landis JR, Koch GG.** The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.
10. **Cohen J.** A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37-46
11. **Fleiss JL, Slakter MJ, Fischman SL, Park MH, Chilton NW.** Inter-examiner reliability in caries trials. J Dent Res. 1979;58:604-9.
12. **Taşkesen A, Demirkale I, Okkaoğlu MC, Özdemir M, Bilgili MG, Altay M.** Intraobserver and interobserver reliability assessment of tibial plateau fracture classification systems. Eklem Hastalik Cerrahisi. 2017;28:177-81.
13. **Doornberg JN, Rademakers MV, van den Bekerom MP. *et al.*** Two-dimensional and three-dimensional computed tomography for the classification and characterization of tibial plateau fractures. Injury. 2011;42:1416-25.
14. **te Stroet MA, Holla M, Biert J, Kampenet AV.** The value of a CT scan compared to plain radiographs for the classification and treatment plan in tibial plateau fractures. Emerg Radiol. 2011;18:279-83.
15. **Brunner A, Horisberger M, Ulmar B, Hoffmann A, Babst R.** Classification systems for tibial plateau fractures; does computed tomography scanning improve their reliability? Injury. 2010;41:173-8.
16. **Hu YL, Ye FG, Ji AY, Qiao GX, Liu H-F.** Three-dimensional computed tomography imaging increases the reliability of classification systems for tibial plateau fractures. Injury. 2009;40:1282-5.
17. **Mellema JJ, Doornberg JN, Molenaars RJ, Ring D, Kloen P.** Traumaplatform Study Collaborative & Science of Variation Group. Interobserver reliability of the Schatzker and Luo classification systems for tibial plateau fractures. Injury. 2016;47:944-9.
18. **ZhuY, YangG, LuoCF, *et al.*** Computed tomography-based Three-Column Classification in tibial plateau fractures: introduction of its utility and assessment of its reproducibility. J Trauma Acute Care Surg. 2012;73:731-7.
19. **Billi AM, Masdeu GM, Bermejo ER, Ibañez M, Gelber PE.** What is the most reproducible classification system to assess tibial plateau fractures? Int Orthop. 2017;41:1251-1256.
20. **Millar SC, Arnold JB, Thewlis D, Fraysse F, Solomon LB.** A systematic literature review of tibial plateau fractures: What classifications are used and how reliable and useful are they? Injury. 2018;49:473-490.
21. **Rasmussen S, Madsen PV, Bennicke K.** Observer variation in the Lauge-Hansen classification of ankle fractures. Precision improved by instruction. Acta Orthop Scand. 1993;64:693-4.