RESEARCH ARTICLE

Ten-Segment Classification has Lowest Inter/Intra-Observer Reliability as Compared to Schatzker, Three-Column and AO Systems for Tibial Plateau Fractures: A Comparison Based on Surgeons' Experience

Nitesh Gahlot, MD, MBBS, MS, DNB, MNAMS; Kishor Kunal, MD; Abhay Elhence, MD; Umesh Meena, MD; Akshat Gupta, MD; Jeshwanth Netaji, MD; Dharampal Swami, MD; Meghal Goyal, MD; Ashraf Jamal, MD

Research performed at All India Institute of Medical Sciences (AIIMS), Jodhpur, India

Received: 6 February 2022 *Accepted: 9 November 2022*

Abstract

Objectives: The primary aim of this study was to assess the reliability of the ten-segment classification system proposed (TSC) by Krause *et al*. and see how it compares with the traditionally used Schatzker classification, AO classification system, and Luo's "Three columns" classification (ThCC) system. The second aim of this study was to assess the inter-observer reliability of the above classifications based on professional experience by comparing the entry level of residents (1 year into postgraduation), senior residents (1 year after postgraduation completion), and faculty (>10 years after postgraduation completion).

Methods: 50 TPFs were classified by a 10-segment classification system, and its intra-observer (at 1-month interval) and inter-observer reproducibility was checked using *k* values by three different groups with varying levels of experience (Group I, II, and III comprised of 2 juniors residents, senior residents and consultants each), and the same was compared for three other common classification systems (Schatzker, AO and 3 –column).

Results: 10-segment classification showed least *k* for both inter-observer (0.08) and intra-observer (0.03) reliability. Highest individual inter-observer (*k*= 0.52) and intra-observer reliability (*k*= 0.31) was for Schatzker classification in Group I. Lowest individual inter-observer and intra-observer reliability was seen for 10-segment classification (*k*= 0.07) and AO classification system (*k*= -0.03) respectively.

Conclusion: 10-segment classification showed the lowest *k* for both inter-observer and intra-observer reliability. The inter-observer reliability for the Schatzker, AO, and 3- column classifications reduced with increasing experience of the observer (JR>SR>Consultant). A possible reason could be a more critical evaluation of the fractures with increasing seniority.

Level of evidence: I

Keywords: Clinical competence,Computer-assisted, Image processing, Observer variation, Tomography, X-Ray computed

Introduction

The incidence of tibial plateau fracture has been reported to be 10.3 per 100,00 annually.[1] Numerous classification systems for proximal tibial plateau fractures have been described based on plain radiographs or CT scans.[1-3] The initial attempt to classify tibial plateau fractures on radiographs goes back to 1939 when the first system was given by Marchant,[2] dividing the TPFs into three types- separated, depressed, and combined. In 1960, the first comprehensive classification was given by Duparc and Ficat (later modified in 1990) based on the fracture

Corresponding Author: Nitesh Gahlot, Department of Orthopaedics, All India Institue of Medical Sciences (AIIMS), Jodhpur, India

Email: doc.nitesh@gmail.com, gahlotn@aiimsjodhpur.edu.in

THE ONLINE VERSION OF THIS ARTICLE ABJS.MUMS.AC.IR

orientation in the axial, coronal and sagittal planes.[3] In 1979, Schatzker *et al.*[4] proposed their classification, dividing the proximal tibia fractures into six types based on radiographs. It is still considered the most widely used system. In 1987, Muller and the AO group proposed the AO classification, which was more extensive and classified TPFs into three types and nine subtypes.[5] While commonly used Schatzker's and AO/Muller's classifications are based on plain radiographs, Luo's three-column classification system is based on the axial section of the CT scan.[4-6] A wide gamut of fracture patterns leads to the failure of these classification systems in properly assisting the planning of anatomic reconstruction. This has led to a more extensive analysis of fracture fragments; hence, the classification moved from radiograph to CT. Krause's ten-segment classification concept[7] has been proposed to describe the fracture fragments on CT comprehensively.

However, it becomes challenging and confusing for the orthopaedic surgeon to decide the best classification system for planning the surgical intervention and communicating with peers simultaneously. Also, it is difficult to reproduce the effects of a surgical intervention planned based on a particular classification system if other surgeons can't reliably reproduce the system above. Thus, the primary objective of this study was to assess the reliability of the ten-segment classification system proposed (TSC) by Krause *et al.*[7] and see how it compares with the conventionally used Schatzker classification,[4] AO classification system[5] and Luo's "Three columns" classification (ThCC) system.[6,8] A secondary objective was to assess the inter-observer reliability of all the classifications above based on the professional experience by comparing the entry-level resident (1 year into postgraduation), senior residents (1 year after postgraduation completion), and faculty (>10 years after postgraduation completion). This shall give insight into the complexity of the classification systems and whether they are simple enough to be interpreted correctly by junior orthopaedic surgeons. To the author's knowledge, this study is the first to assess Krause's classification system for reliability and reproducibility and the first to assess the above classification regarding observers' clinical experience.

Materials and Methods

Patients who sustained TPFs presented to our tertiary care centre and met the inclusion criteria were included in the study. The study period was from 2016 to 2020. The clinical details, radiographs, and CT images were collected from the hospital archives. Inclusion criteria were: intra-articular fractures of the tibial plateau and skeletally mature patients in the age group of 16 - 65 years. Patients with congenital anomalies of the affected limb or residual effects of neurological disorders such as poliomyelitis, pathological proximal tibia fractures, and pre-existing knee deformity were excluded from the study.

The blinded radiological data of the proximal tibia fracture patients were collected. The radiology included radiographs of the knee with the proximal tibia in anteroposterior and lateral views and 2D & 3D computed tomography (CT) scans [Figure 1,2]. The observers participating in the study were given the original studies published for all four classifications for their understanding [Figure 3-5]. After that, they were shown the radiology of the fractures randomly. An independent observer selected and compiled the relevant images for evaluation.

Figure 1. Anteroposterior and lateral radiograph of a patient with a proximal tibia fracture. Radiographs were used while classifying according to the Schatzker classification

Figure 2. CT scan images of the same patient. (A) 3D reconstruction of the fracture. (B) Axial, coronal, and sagittal cuts of the fracture are used for classifying according to the AO, three-column, and ten-column classifications

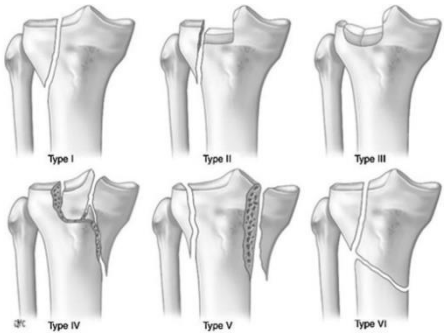

Figure 3. Schatzker classification

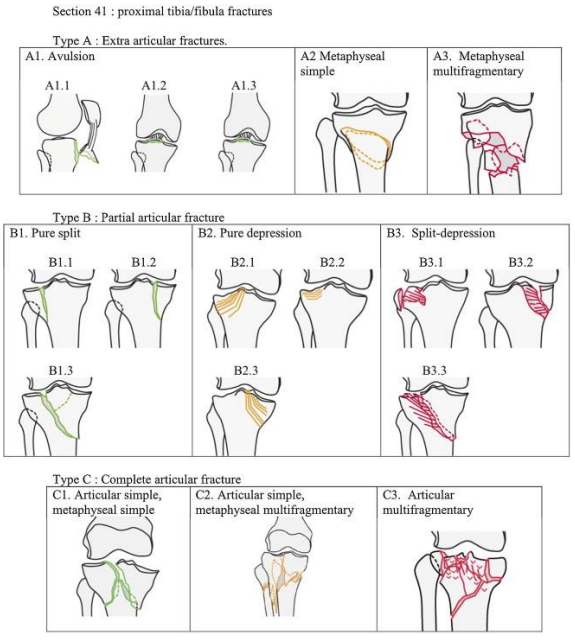

Figure 4. AO classification

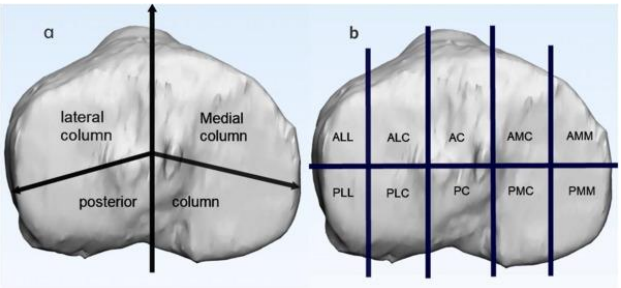

Figure 5. Three-column and ten-segment classification

The observers were divided into three groups based on their clinical experience: Group I (two junior residents, one year into the post-graduation), Group II (senior resident, two years after completing the post-graduation), and Group III (Faculty, >7 years after post-graduation). Each observer classified the fracture twice at the two-week interval, with a randomly changed sequence of radiographs. The response was noted for analysis.

The statistical analysis was done on the SPSS 25 (Armonk, NY: IBM Corp., 2017) software. The k (Kappa) value was used to calculate the reliability of the fracture classification system.

k is a coefficient of agreement that varies from +1 (perfect agreement) to '0' (agreement no better than chance) to –1 (representing absolute disagreement). Inter-observer, as well as the intra-observer reliability of the TSC, [6] was noted and compared with the same for Schatzker, ThCC, and the AO system of classification.[4-6,8] Guidelines proposed by Landis and Koch shall be used to grade the different levels of reliability based upon k values- k> 0.81 as perfect, k= 0.61-0.8 as substantial, k= 0.41-0.6 as moderate, k= 0.21-0.4 as fair and k< 0.2 as slight correlation.

Results

The mean age of patients was 39.56 (+/-13.27) years. There were 46 males and 4 females. The mean age of males was 39.74 (+/-12.51) years, while those of females were 37.5 (+/-22.93) years. Out of 50 knees, 26 were on the right side, and 24 were on the left. Table 1 gives the results of *k* statistical analysis of inter-observer and intra-observer reliability [Table 1].The range of *k* for *Schatzker classification* was from 0.23- 0.69 for inter-observer and 0.21-0.33 for intra-observer reliability, with an average of 0.45 and 0.29, for inter-observer and intra-observer reliability, respectively [Table 2].

Table 1.Comparison of Schatzker, 3-column, AO and 10-segment classification systems among various observers' terms of mean inter-observer and intra-observer reliability. (JR: junior resident; SR: senior resident; C: consultant). The value in bracket shows p-value. Column - "self"- shows intra-observer reliability (k)

Schatzker	JR1	JR2	SR1	SR2	C1	C2	Self
JR1	-	0.69 (<0.01)	0.51 (<0.01)	0.55 (<0.01)	0.59 (<0.01)	0.32 (<0.01)	0.33 (<0.01)
JR2	-	-	0.54 (<0.01)	0.48 (<0.01)	0.47 (<0.01)	0.31 (<0.01)	0.28 (<0.01)
SR1	-	-	-	0.34 (<0.01)	0.43 (<0.01)	0.37 (<0.01)	0.31 (<0.01)
SR2	-	-	-	-	0.57 (<0.01)	0.24 (<0.01)	0.27 (<0.01)
C1	-	-	-	-	-	0.23 (<0.01)	0.33 (<0.01)
C2	-	-	-	-	-	-	0.21 (<0.01)
Average	0.53	0.50	0.44	0.44	0.46	0.29	
3-column	JR1	JR2	SR1	SR2	C1	C2	Self
JR1	-	0.48 (<0.01)	0.28 (<0.01)	0.18 (.05)	0.51 (<0.01)	0.1 (0.22)	0.24 (0.01)
JR2	-	-	0.17 (0.08)	0.29 (<0.01)	0.49 (<0.01)	0.08 (0.29)	0.04 (0.66)
SR1	-	-	-	0.47 (<0.01)	0.22 (0.04)	0.06 (0.32)	0.37 (<0.01)
SR2	-	-	-	-	0.33 (<0.01)	0.01 (0.91)	0.12 (0.21)
C1	-	-	-	-	-	0.15 (0.03)	0.28 (<0.01)
C2	-	-	-	-	-	-	0.05 (0.41)
AO	JR1	JR2	SR1	SR2	C1	C2	Self
JR1	-	0.47 (<0.01)	0.58 (<0.01)	0.55 (<0.01)	0.48 (<0.01)	0.24 (<0.01)	-0.05 (0.43)
JR2	-	-	0.42 (<0.01)	0.49 (<0.01)	0.34 (<0.01)	0.15 (0.02)	-0.01 (0.94)
SR1	-	-	-	0.40 (<0.01)	0.34 (<0.01)	0.13 (0.04)	0.05 (0.51)
SR2	-	-	-	-	0.50 (<0.01)	0.25 (<0.01)	0.07 (0.31)
C1	-	-	-	-	-	0.13 (0.04)	0.02 (0.76)
C2	-	-	-	-	-	-	0.10 (0.09)
Average	0.46	0.37	0.37	0.44	0.36	0.18	
10-segment	JR1	JR2	SR1	SR2	C1	C2	Self
JR1	-	0.07 (0.21)	0.07 (0.18)	0.02 (0.63)	-0.06 (0.16)	0.03 (0.56)	0.12 (0.11)
JR2	-	-	0.07 (0.29)	0.06 (0.25)	0.14 (0.01)	0.14 (0.02)	-0.03 (0.55)
SR1	-	-	-	0.13 (0.02)	0.18 (<0.01)	0.08 (0.20)	0.08 (0.11)
SR2	-	-	-	-	0.16 (<0.01)	0.03 (0.48)	0.01 (0.93)
C1	-	-	-	-	-	0.05 (0.40)	0.07 (0.18)
C2	-	-	-	-	-	-	-0.08 (0.85)
Average	0.03	0.10	0.11	0.08	0.09	0.07	

Table 2. Comparison of various classification systems in terms of mean inter-observer and intra-observer reliability (JR: Junior Resident, SR: Senior Resident). The boxes marked in yellow shows highest inter-observer k values among group

Classification	Overall		JR (Group 1)		SR (Group 2)		Consultant (Group 3)	
	Inter-observer	Intra-observer	Inter-observer	Intra-observer	Inter-observer	Intra-observer	Inter-observer	Intra-observer
Schatzker	0.45	0.29	0.52	0.31	0.44	0.29	0.38	0.27
AO	0.37	0.03	0.42	-0.03	0.41	0.06	0.27	0.06
3 column	0.26	0.19	0.31	0.14	0.25	0.25	0.21	0.17
10-segment	0.08	0.03	0.07	0.04	0.10	0.05	0.08	-0.01

The k for AO classification ranged from 0.13-0.58 for inter-observer and -0.05-0.10 for intra-observer reliability, with an average of 0.37 and 0.03, respectively. The k for ThCC ranged from 0.01-0.49 for inter-observer and 0.04-0.37 for intra-observer reliability, with an average of 0.26 and 0.19, respectively. The range of k for TSC was from -0.06-0.18 for inter-observer and -0.08-0.12 for intra-observer reliability, with an average of 0.08 and 0.03, respectively.

For group-wise reliability, the highest mean inter-observer k was observed for the Schatzker classification system (0.52 and the lowest for the 10-segment classification system (0.07), both in the JR group. The highest mean intra-observer k was observed for the Schatzker classification system (0.31) and the lowest for both AO (-0.03) both in the JR group [Table 2]. The 10-segment classification system can produce over a thousand combinations; thus, the number of patients ideally shall be higher to measure a pure correlation. The 10-segment system was reduced to numbers rather than unique sets to determine reliability to achieve assessable data.

Discussion

With the advent of computerized tomography (CT), a better understanding of fracture anatomy became possible, leading to the introduction of the column concept for TPFs. The first such classification was the three-column concept proposed by Luo et al. in 2010 and later updated by Wang et al. in 2016.[6,8] According to them, Schatzker and the older classification systems were two-dimensional classifications based mainly on the anteroposterior radiographs of the proximal tibia. As such, they could not identify and adequately classify posterior shear fractures and coronal split of the posterior column of the tibial plateau. The three columns proposed were: medial, lateral, and posterior, and "one independent articular depression with a break of the column wall was defined as a fracture of the relevant column".[6,8,9] Recently in 2014, Chang et al. proposed a modification of the ThCC in which the tibial plateau was broken down into four columns instead of three: anteromedial, anterolateral, posteromedial, and posterolateral columns.[10] The latest scheme of TPF classification is that given by Krause et al.[7] In this system, the proximal 3 cm of the tibial plateau was first broken down into an anterior and a posterior column in the axial plane. The frontal plane was divided into central, medial, and lateral portions. The latter two were further subdivided into a lateral and central segment for the lateral plateau and a medial and central segment for the medial plateau. This comprehensive system can identify most fracture patterns that would have been missed even by the three/four column classification.

However, some concerns regarding the TSC system were raised by Dhillon *et al.*[11] They felt that the classification was too cumbersome to be used in routine practice and that the three/four column classification was better suited for TPFs. However, there was no supporting data. This study validates the poor reliability of TSC. Multiple authors have studied the reliability of AO, Schatzker, and ThCC systems, but none have assessed the TSC system. Given the comprehensive nature of these classifications, it becomes important to determine how easily they can be understood and applied practically by the residents as they are usually the first point of contact in the emergency and are responsible for communicating with the faculty for planning and operating the patient.

Charalambous *et al.*[12] showed inter-observer reliability for Schatzker and AO to be 0.41 and 0.43, respectively, while intra-observer to be 0.57 and 0.53. Similar inter-observer reliability was observed in our study. However, intra-observer reliability was relatively low. Maripuri *et al.*[13] showed inter-observer reliability for Schatzker and AO to be 0.47 and 0.36, respectively, while intra-observer to be 0.93 and 0.88. The highest inter-observer reliability for AO was noted to be 0.43 by Charalambous *et al.*[12] and the lowest of 0.357 by Giquel *et al.*[14] For Schatzker, inter-observer reliability has been reported from 0.38-0.47.[14, 15] Millan-Billi[16] reported higher reproducibility from 0.62 of AO to 0.73 for 3-column classification between observers. In our study, the highest inter-observer correlation (*k* 0.69) was observed between two junior residents (group 1) for the Schatzker classification, with most inter-observer reliabilities being above 0.4 in the same group. The highest recorded *k* value reduced with AO (0.58) followed by 3-column classification (0.49) and then was least for the 10-segment classification system (0.18). 10-segment classification system achieved both the lowest inter-observer and intra-observer kappa in our study [Table 2]. During the present study, multiple times, the observers classified different columns in Krouse's TSC for the same fracture, with the total number of segments being the same. As we could not find any method in the classification to distinguish them, for analysis, the groups were made based on the number of segments.

On comparing the reliability of the classifications inside the three groups of observers individually, there was the highest intra-observer reliability with the Schatzker classification, followed by AO, 3-column, and 10-segment classifications in that order. The overall intra-observer reliability was highest with the Schatzker classification, followed by ThCC, AO, and Krouse's TSC. This trend was also present inside the individual groups except for group 1, where the AO system fared worse than the ten segments. This indicates better

reliability of the classification system with fewer subclasses.

The inter-observer reliability for the Schatzker, AO, and 3-column classifications reduced with increasing observer experience (JR>SR>Consultant), which goes opposite to the usual assumption of improvement with seniority, i.e., with increasing experience, the reproducibility of a classification decreases. The probable cause of this might be a better critical fracture analysis with more clinical experience. However, this trend was not seen in the 10-column classification, which remained at low interobserver reliability overall and in all three groups (slightly higher in the SR group) [Table 2]. Ours is the first study to evaluate these classifications regarding the observers' experience. We consider this an essential aspect as it indicates the ease of understanding and complexity of the classification. A more straightforward classification is also essential and well understood by junior orthopaedic surgeons.

In places with adequate facilities where getting a CT scan is not a limitation, ThCC shall prove useful over the TSC system in terms of reliability. Different studies have mentioned the usefulness of the 3- column and 10-segment classification in preoperative planning of the TPFs.[9,17] The AO and Schatzker classification systems were proposed on plain radiographs with limited ability to assess the 3D anatomy; there are fewer chances of having a variable opinion among observers, resulting in improved inter-observer and intra-observer reliability. The ten-segment classification considers the fractures seen in the proximal 3 cm of the tibia; we believe this needs slight modification. We recommend that the classification consider the column pillar broken if there is a continuity breach of the cortex anywhere, no matter how far down the metaphysis. Few TPFs are unique and challenging to classify, e.g., the posterior shear fractures with posterior subluxation of the femur can be classified in AO B1 and as a posterior column in 3-column and 10-segment classification, but not under any group of Schatzker classification.

In the present study, we found that the Schatzker classification system had better inter-observer and intra-observer reliability than the other three. Anwar *et al.*[18] proposed a two-column classification system and showed excellent reliability with a kappa >0.9. *Reducing the number of columns reduces the number of combinations, thus increasing the likelihood of being similar between two observers.* Maripuri *et al.*[13] has recommended a need for the development of a comprehensive approach to judging the quality of fracture classification systems.

The limitation of the present study includes the limited sample size for calculation of reliability in the Ten-segment classification system; we have tried to compensate for that by increasing the number of observers. To the authors' knowledge, this is the first study to assess the newer 10-segment classification for reliability and compare it with the established classification systems. We are also the first to determine the four major classification systems for TPFs for their utility and ease by orthopedic surgeons with variable clinical experience. The data presented is extensive in the study, but we believe it is necessary; as the TCS is CT based, we needed to compare it with a similar CT-based classification (ThCC) given that the AO and Schatzker systems are X-ray based.

Conclusion

10-segment classification showed the lowest k for both inter-observer and intra-observer reliability compared. The inter-observer reliability for the Schatzker, AO, and 3-column classifications reduced with increasing experience of the observer (JR>SR>Consultant). A possible reason could be a more critical evaluation of the fractures with increasing seniority.

Acknowledgement

Not applicable

Conflict of interest: The authors declare that they have no competing interests.

Funding: Not applicable

Ethics approval and consent to participate: Not applicable

Nitesh Gahlot MD [1]
Kishor Kunal MD [1]
Abhay Elhence MD [1]
Umesh Meena MD [2]
Akshat Gupta MD [1]
Jeshwanth Netaji MD [1]
Dharampal Swami MD [1]
Meghal Goyal MD [1]
Ashraf Jamal MD [1]

1 Department of Orthopaedics, All India Institute of Medical Sciences (AIIMS), Jodhpur, India
2 Department of Orthopaedics, SMS Medical College, Jaipur, Rajasthan, India

References

1. Elsoe R, Larsen P, Nielsen NP, Swenne J, Rasmussen S, Ostgaard SE. Population-Based Epidemiology of Tibial Plateau Fractures. Orthopedics. 2015; 38(9):e780-6. doi:10.3928/01477447-20150902-55
2. Gerard-Marchant P. Fractures des plateaux tibiaux. Rev Chir Orthop. 2002; 26:499-546.
3. Huten D, Duparc J, Cavagna R. Fracture's récentes des plateaux tibiaux de l'adulte. Editions techniques, EMC (Paris, France), Appareil locomoteur. 1990; 14082:A10.
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. Müller ME, Nazarian S, Koch P, eds. Classification AO des fractures: 1: Les os longs. French Edition. Springer-Verlag; 1987.
6. Luo CF, Sun H, Zhang B, Zeng BF. Three-column fixation for complex tibial plateau fractures. J Orthop Trauma. 2010; 24(11):683-92. doi:10.1097/BOT.0b013e3181d436f3
7. Krause M, Preiss A, Muller G, et al. Intra-articular tibial plateau

fracture characteristics according to the "Ten segment classification". Injury. 2016; 47(11):2551-2557. doi:10.1016/j.injury.2016.09.014
8. Wang Y, Luo C, Zhu Y, et al. Updated Three-Column Concept in surgical treatment for tibial plateau fractures - A prospective cohort study of 287 patients. Injury. 2016; 47(7):1488-96. doi:10.1016/j.injury.2016.04.026
9. Zhu Y, Yang G, Luo CF, 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(3):731-7. doi:10.1097/TA.0b013e31825c17e7
10. Chang SM, Hu SJ, Zhang YQ, et al. A surgical protocol for bicondylar four-quadrant tibial plateau fractures. Int Orthop. 2014; 38(12):2559-64. doi:10.1007/s00264-014-2487-7
11. Dhillon MS, Patel S, Puneeth K. Simple four column classification can dictate treatment for intra articular tibial plateau fractures much better than ten segment classification. Injury. 2017; 48(6):1276-1278. doi:10.1016/j.injury.2017.03.031
12. Charalambous CP, Tryfonidis M, Alvi F, et al. Inter- and intra-observer variation of the Schatzker and AO/OTA classifications of tibial plateau fractures and a proposal of a new classification system. Ann R Coll Surg Engl. 2007; 89(4):400-4. doi:10.1308/003588407X187667
13. Maripuri SN, Rao P, Manoj-Thomas A, Mohanty K. The classification systems for tibial plateau fractures: how reliable are they? Injury. 2008; 39(10):1216-21. doi:10.1016/j.injury.2008.01.023
14. Gicquel T, Najihi N, Vendeuvre T, Teyssedou S, Gayet LE, Huten D. Tibial plateau fractures: reproducibility of three classifications (Schatzker, AO, Duparc) and a revised Duparc classification. Orthop Traumatol Surg Res. 2013; 99(7):805-16. doi:10.1016/j.otsr.2013.06.007
15. Walton NP, Harish S, Roberts C, Blundell C. AO or Schatzker? How reliable is classification of tibial plateau fractures? Arch Orthop Trauma Surg. 2003; 123(8):396-8. doi:10.1007/s00402-003-0573-1
16. Millan-Billi A, Gomez-Masdeu M, Ramirez-Bermejo E, Ibanez M, Gelber PE. What is the most reproducible classification system to assess tibial plateau fractures? Int Orthop. 2017; 41(6):1251-1256. doi:10.1007/s00264-017-3462-x
17. Tuladhar R, Hu SJ, Chang SM. The articular and non-articular proportion in segment classification of tibial plateau fractures. Injury. 2019; 50(3):818-819. doi:10.1016/j.injury.2019.01.021
18. Anwar A, Zhang Y, Zhao Z, et al. Two column classification of tibial plateau fractures; description, clinical application and reliability. Injury. 2019; 50(6):1247-1255. doi:10.1016/j.injury.2019.04.021