INTER- AND INTRA-OBSERVER RELIABILITY OF SCHATZKER, AO, AND LUO CLASSIFICATIONS FOR TIBIAL PLATEAU FRACTURES

AVALIAÇÃO INTER E INTRAOBSERVADORES DAS CLASSIFICAÇÕES DE SCHATZKER, AO E LUO PARA FRATURAS DO PLATÔ TIBIAL

JONATAS BRITO DE ALENCAR NETO[1], CLODOALDO JOSÉ DUARTE DE SOUZA[1], PEDRO RAFAEL REIS COELHO[1], PEDRO GOMES DE FREITAS JÚNIOR[1], MÁRCIO BEZERRA GADELHA LOPES[2], FRANCISCO JOSÉ MAIA PINTO[3]

1. Instituto Dr. José Frota, Orthopedics and Traumatology Service, Fortaleza, CE, Brazil.
2. Hospital Geral de Fortaleza, Orthopedics and Traumatology Service, Fortaleza, CE, Brazil.
3. Universidade Estadual do Ceará, School of Medicine, Fortaleza, Ceará, Brazil.

ABSTRACT

Objective: To verify inter- and intra-observer agreement of three classification systems for tibial plateau fractures – Schatzker, AO/ASIF, and Luo's – among orthopedic surgery residents. Methods: This cross-sectional study was conducted with 29 observers. Radiographic and tomographic imaging of the knee of 15 patients presenting with fractures were evaluated. After six weeks, the test was reapplied. The level of agreement was calculated by the Kappa index. Results: In test 1, inter-observer agreement of all residents, according to the Kappa index, for Schatzker classification was 0.226, for AO 0.431, and Luo's 0.319. In test 2, values were 0.316, 0.333, and 0.347, respectively ($p < 0.001$). Regarding intra-observer analysis, the mean Kappa indexes of 1st-year residents were: Schatzker, 0.20; AO, 0.32; and Luo's, 0.3. For 2nd-year residents, means were: 0.51, 0.58, and 0.38, respectively. As for 3rd-year, results were 0.42, 0.42, and 0.41, respectively ($p < 0.001$). Conclusion: AO/ASIF showed a better reproducibility than other classifications, with substantial inter- and intra-observer agreement. We also found a stronger agreement among 2nd- and 3rd-year residents. ***Level of Evidence III, Diagnostic Studies – Investigating a Diagnostic Test.***

Keywords: Tibial Fractures. Classification. Correlation Analyses.

RESUMO

Objetivo: Verificar a concordância interobservador e intraobservador de três classificações utilizadas nas fraturas do platô tibial: Schatzker, AO/ASIF e a de Luo entre residentes de ortopedia. Métodos: estudo transversal, por conveniência, com 29 observadores. Foram avaliadas imagens radiográficas de joelho e tomografia computadorizada em 15 pacientes com fraturas de platô tibial, submetidos a tratamento cirúrgico, sendo reaplicado o teste após seis semanas. O nível de concordância foi calculado através do índice Kappa. O estudo foi submetido ao Comitê de Ética em Pesquisa em humanos da instituição. Resultados: A concordância interobservador dos residentes dos três anos juntos de acordo com o índice Kappa para a classificação Schatzker, AO e tomográfica no teste 1 foi 0,226, 0,431 e 0,319 respectivamente e no teste 2 0,316, 0,333, 0,347 respectivamente. Na análise intraobservador, foi calculada a média dos índices Kappa dos residentes do 1º ano: classificação Schatzker, 0,20; AO 0,32; e tomográfica 0,3. Já a média para os residentes do 2º ano Schatzker foi 0,51, 0,58 e 0,38, respectivamente. Os resultados para os residentes do 3º ano foram 0,42, 0,42 e 0,41 respectivamente. Conclusão: A classificação AO/ASIF mostrou-se ter uma melhor reprodutibilidade em relação as outras com uma concordância substancial na análise interobservador e intraobservador. Evidenciou-se também que os residentes do 2º e 3º ano tem uma concordância maior. ***Nível de Evidência III, Estudos diagnósticos – Investigação de um exame para diagnóstico.***

Descritores: Fraturas da Tíbia. Classificação. Análises de Correlação.

INTRODUCTION

Tibial plateau fractures account for 1 to 2% of all fractures and involve the articular surface of the proximal tibia.[1,2] Such trauma has a bimodal distribution, affecting mainly young adults and individuals between the fourth and sixth decades of life, and may be sustained in high- or low-energy mechanisms.[3,4] They result from axial compression associated with varus or valgus deviation, which may cause depression or shearing of the load-bearing area.[1,5]

This type of fracture requires a deep understanding of the lesion pattern and, consequently, classification systems for surgical planning and prognosis estimates.[1] These evaluations often entails radiographs, especially in the Schatzker classification, the most used method worldwide.[6]

Although many possible classification systems for tibial plateau fractures are available, the most employed are: Schatzker, the Association for Osteosynthesis/Association for the

Study of Internal Fixation (AO/ASIF) classification, and Luo's tomographic classification.

The Schatzker classification is based on anteroposterior plain radiographs and divides injuries into six types, ranging from grade 1 to 6.[7-9] In turn, AO/ASIF classifies fractures according to the alphanumeric code 41 (4 – tibia, 1 – proximal region) into the following standards: A (non-articular), B (partial articular), and C (total articular), whereby the higher the standard, the greater the severity and energy of the injury.[10] Luo's classification draws on tomography sections to classify fractures according to their localization within the axial plane, subdividing them into fractures of the lateral, medial, and posterior column. These subgroups may eventually intersect.[11]

When selecting the ideal treatment, the classification of fractures gains increasing relevance. Schatzker classification, still important in the field of tibial plateau fractures, provides a definition of the pathological anatomy and suggests therapeutic strategies. The AO/ASIF classification has an adequate performance for the proximal third of the tibia and remains the leading international method for classifying fractures. Luo's tomographic classification emerged as an alternative method for visualizing the fracture in three dimensions and defining appropriate treatment strategies.

This study aimed to evaluate the levels of intra- and inter-observer reliability for the Schatzker, AO/ASIF, and Luo's classification systems in a group of resident physicians in different years of their residency at a tertiary-care trauma hospital to reproduce the behavior of different surgeons from heterogeneous groups.

MATERIALS AND METHODS

A total of 15 patients presenting with tibial plateau fractures who attended to the traumatology service of a level-I trauma hospital from April 2016 to June 2016 were selected by non-probability, convenience, and consecutive sampling.

In that period, 29 observers (resident physicians, of which nine were from the 1st year, ten from the 2nd, and ten from the 3rd) independently and randomly classified radiographs (anteroposterior and profile) and computed tomography (CT) scans of the knee, with 0.6 mm axial slices. A Philips® digital radiography system and a Siemens CT Scan machine with 16 multi-slice channels for sections width of 0.6 mm were used. Images were obtained by photographs taken with an iPhone® 6 with 8-megapixel resolution, and displayed using PowerPoint® 2016 on an Epson overhead projector. After six weeks, the test was reapplied. The level of agreement was calculated by the Kappa index for intra- and inter-observer reliability and interpreted according to Landis and Koch.

All observers had access to the same images and categorized the fractures according to Schatzker, AO/ASIF, and Luo's classification systems. Data were recorded on a specific paper questionnaire.

Before each application, an orthopedist specialized in knee trauma briefly reviewed the three classifications used. An illustration of each classification with subtitles was also provided to all observers.

After completion, results were collected and analyzed using the Statistical Package for Social Sciences – SPSS® v. 23.0 (Chicago, USA) to determine Kappa coefficient. Table 1 shows how intra- and inter-observer agreement was determined. Statistical analysis was performed using Cohen's Kappa (k) statistic to evaluate inter-observer reliability between tests, with 95% confidence interval (95%CI). When analyzing agreement among residents (1st, 2nd, and 3rd years), Fleiss' Kappa coefficient of agreement was adopted, with 95% confidence interval (95%CI) and tests to assess whether the coefficient was statistically different from zero. The tested hypothesis was: agreement (ρ) equal to zero indicated no inter-observer agreement, Kappa statistic equals 1 indicated perfect agreement, and close to or less than zero indicated disagreement, that is, the observed agreement is less suitable than by chance. This study was conducted following the basic principles for ethical researches involving humans – as autonomy of research participants, justness, beneficence, and non-maleficence, – according to Resolution 466/12. Participants agreed to participate by signing a consent form prior to the pilot test, usability test, and application effectiveness test.

The study was submitted to the Human Research Ethics Committee (CEP) and approved under protocol No. 2,087,850.

RESULTS

Among 1st-year residents, the results of tests 1 and 2 showed values close to the kappa coefficients in the Schatzker (0.125 and 0.114) and AO/ASIF (0.217 and 0.220) classifications. However, Luo's tomographic classification showed a higher value in test 1 (0.264 and 0.213). According to Landis and Koch, agreement for the Schatzker classification is considered slight, and for the AO and Luo's tomographic classifications it is fair.

As for 2nd-year residents, values were close to the kappa coefficients both in Luo's tomographic classification (0.390 and 0.364) and Schatzker (0.246 and 0.380). In test 2, AO/ASIF (0.453 and 0.538) showed higher coefficients values. According to Landis and Koch, the agreement in Schatzker and Luo's classifications was fair, and in AO/ASIF it was moderate.

Regarding 3rd-year residents, the kappa coefficient increased from test 1 to test 2 in Schatzker (0.226 and 0.316) and Luo's (0.319 and 0.347), but decreased in AO (0.431 and 0.333). For Landis and Koch, most agreements were fair, except in AO for test 1, in which it was moderate.

In the three classification systems, Fleiss' kappa coefficients of agreement were lower in test 1 for 1st-year residents and higher for 2nd-year residents. AO/ASIF presented the highest values, which, according to Landis and Koch, was considered a moderate agreement. As for test 2, Fleiss' kappa coefficients of agreement were also lower for 1st-year residents and higher for 2nd-year residents in all three classifications. AO/ASIL presented the highest observed value, which, according to Landis and Koch, was considered a moderate agreement (Table 1).

The mean kappa value from the two tests applied, calculated from the intra-observer analysis of 1st-year residents, was 0.25 for Schatzker, 0.32 for AO/ASIF, and 0.30, for Luo'stomographic classification. We found a stronger agreement in the AO/ASIF classification among 1st-year residents.

We performed an agreement analysis among 1st-year residents for tests 1 and 2. Resident 5 had the lowest kappa coefficient, in Schatzker classification (0.077), and resident 9 had the highest, in the AO/ASIF classification (1.000). In the Schatzker classification, resident 1 obtained 0.429, resident 6 obtained 0.634, and resident 9 0.511, considered moderate. As for kappa values in AO/ASIF, resident 3 obtained 0.405, resident 4 0.496, resident 8 0.455, and resident 9 1.0. In turn, kappa values in the Luo's tomographic classification were 0.420 for resident 3, 0.400 for resident 4, 0.457 for resident 6, and 0.630 for resident (Figure 1).

Table 1. Fleiss' kappa coefficients (K) of inter-observer agreement, with 95% confidence intervals (95%CI), and p-values according to resident group, test, and classification system of patients treated at the tertiary hospital from April to June 2016.

Resident	Test	Aspect	K	95%CI		p-value
1st year	1	Schatzker	0.125	0.086	0.164	< 0.001
		AO	0.217	0.147	0.288	< 0.001
		Tomographic	0.264	0.221	0.307	< 0.001
	2	Schatzker	0.114	0.072	0.157	< 0.001
		AO	0.220	0.148	0.293	< 0.001
		Tomographic	0.213	0.172	0.254	< 0.001
2nd year	1	Schatzker	0.246	0.206	0.285	< 0.001
		AO	0.453	0.385	0.522	< 0.001
		Tomographic	0.390	0.348	0.433	< 0.001
	2	Schatzker	0.380	0.339	0.422	< 0.001
		AO	0.538	0.469	0.607	< 0.001
		Tomographic	0.364	0.321	0.408	< 0.001
3rd year	1	Schatzker	0.226	0.187	0.266	< 0.001
		AO	0.431	0.354	0.508	< 0.001
		Tomographic	0.319	0.276	0.362	< 0.001
	2	Schatzker	0.316	0.275	0.358	< 0.001
		AO	0.333	0.262	0.403	< 0.001
		Tomographic	0.347	0.304	0.389	< 0.001
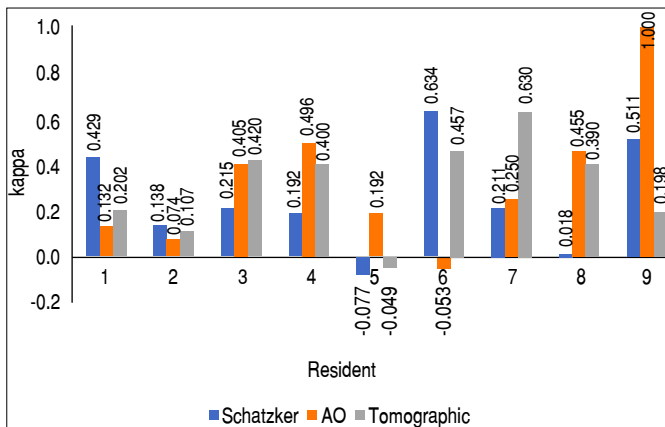
Figure 1. Cohen's kappa coefficients of agreement (k) between tests in 1st-year residents according to classification system and resident.
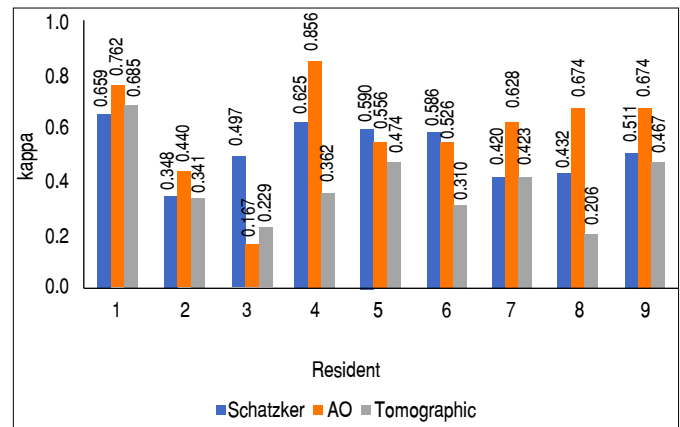
Figure 2. Cohen's kappa coefficients of agreement (k) between tests in 2nd-year residents according to classification system and resident.

The mean kappa value from the two tests applied, calculated from intra-observer analysis among 2nd-year residents, was 0.518 in the Schatzker classification, 0.587 in the AO classification, and 0.306 in the Luo's tomographic classification. We found a more significant agreement in the AO/ASIF classification.

Considering test 1 and 2, kappa coefficients varied less among 2nd-year residents than among 1st-year residents in the three classifications (Figure 2). Schatzker and AO classifications had the least pronounced difference. The lowest value of the kappa coefficient of agreement (0.167), as well as the highest (0.856) occurred in the AO classification for residents 3 and 4. The following residents were classified as moderate or substantial agreement: 1, 5, 7, and 9, in the three classifications; 4, 6, and 8, in the Schatzker and AO classifications; resident 2, in the AO classification, and resident 3, in the Schatzker classification.

The mean kappa value from the two tests applied, calculated in the intra-observer analysis among 3rd-year residents, was 0.426 in the Schatzker classification, 0.426 in the AO/ASIF classification, and 0.415 in the Luo's tomographic classification. We found a stronger agreement in Schatzker and AO classifications among 3rd-year residents, and a smaller difference between absolute kappa values in the three classifications.

Figure 3 shows the differences in the kappa coefficients between tests 1 and 2, both among 3rd- and 1st-year residents, in the three classifications. Resident 7 had the lowest kappa coefficient, in the Schatzker classification (0.052), and resident 1 had the highest in the AO/ASIF classification (0.872). The following residents were classified as moderate or substantial agreement: 1 and 8, in the three classifications; 4 in the Schatzker and AO classifications; 7 in the AO and Luo's tomographic classifications; 3 and 5 in the Luo's tomographic classification; resident 6 in the Schatzker classification, and resident 9 in the AO classification.
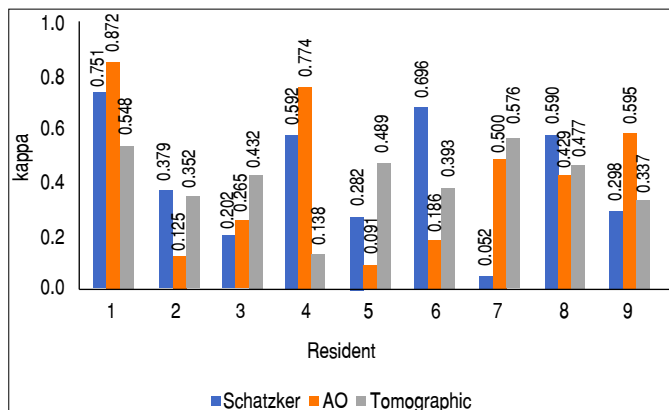
Figure 3. Cohen's kappa coefficients of agreement (k) between tests in 3rd-year residents according to classification system and resident.

DISCUSSION

Classification systems in the fields of orthopedics and traumatology play a key role in the treatment of observed injuries. They assist in standardizing an international classification of lesions, in defining treatment strategies (be it operative or nonoperative), and in predicting outcomes related to the types of fractures presented. When managing a fracture, such criteria must be rigorously fulfilled to obtain a better diagnosis, treatment, and prognosis. For this purpose, our study aimed to determine which classification system meets all requirements, facilitating the surgeon's management of the case.

According to Dirschl et al.,[12] classifications for of tibial plateau fractures present a low inter-observer reproducibility. In their study, 17 observers evaluated 56 fractures based on radiographs, 38 of which also had computed tomography scans. They found a large variability within the reliability of the assessments, as measured by kappa statistics. The higher kappa index was attributable to determining fracture location (0.68), while fracture stability (0.37) and energy (0.29) showed lower reliability. Our study was more focused in analyzing tibial plateau fractures. Yet, rigorous and specific evaluations show a lower inter-observer agreement, indicating a wide variation among evaluated criteria for classifying these fractures.

Walton et al.[9] evaluated 53 knee radiographs in patients presenting with tibial plateau fractures. In comparing Schatzker and AO/ASIF classifications by inter-observer analysis, the authors concluded that AO/ASIF was more reproducible than Schatzker, corroborating our findings. They also verified that both classifications were originally based on radiographic studies.

Mandarino et al.[10] adopted a method similar to ours: they selected 20 examiners to evaluate 20 cases of tibial plateau fractures, in a single application test, using the Schatzker classification. Their study group was composed of 10 orthopedic residents, 5 knee specialists, and 5 general orthopedists. By performing an inter-observer analysis and obtaining a kappa agreement of 0.526, the authors concluded that the Schatzker classification is moderately reproducible, even after grouping them into more homogeneous groups, as showed in our study. As 50% of Mandarino et al.[10] study group consisted of graduated orthopedists, a strong inter-observer agreement (kappa between 0.61 and 0.8) was expected, which was not reflected in the study results. Martin et al.[13] analyzed the inter-observer reproducibility of the AO/ASIF classification in 56 radiographs of tibial plateau fractures, concluding that this classification system is reproducible within this type of fractures, and that the level of agreement in conventional radiographic evaluation depends on the observer's experience. In our study, 2nd-year residents showed a stronger agreement than 3rd-year residents, diverging from the results found by Martin et al. However, we also found that 2nd- and 3rd-year residents showed a stronger agreement than 1st-year residents, corroborating Martin's survey.

Our study poses some limitations, as the difficulty faced in interpreting data of the categorical variables in the Luo's tomographic classification, given they are qualitative variables. The kappa index is more easily applied into quantitative and ordinal (rather than nominal) variables, so qualitative variables demand transforming them into quantitative to enable a better organization.

Another limitation was the limited number of plateau fractures evaluated (n = 15) and observers submitted to the questionnaires (n = 29). Also, the values obtained were not compared with those of orthopedists, knee specialists, or even orthopedic trauma specialists, which could have influenced the results.

We suggest further researchers to proceed with the rigorous investigation on the applicability of existing classifications for tibial plateau fractures, extending the application of questionnaires to general orthopedists and in knee or orthopedic trauma specialists. We also recommend testing the applicability of the classification system that showed greater inter- and intra-observer reproducibility (AO/ASIF) to verify outcomes and complications related to the treatment of these types of fractures. It is also necessary to perform a thoroughly review of the concepts established by the Schatzker classification by testing its applicability and reliability among the evaluated sample, as a way of improving research on tibial plateau fractures.

CONCLUSION

Of the three analyzed classification system, AO/ASIF showed greater intra-observer and inter-observer reproducibility in evaluating tibial plateau fractures, with greater kappa indexes in most of the performed analyses.

ACKNOWLEDGEMENTS

REFERENCES

1. Tomislav C, Cerovec T. Tibial plateau fracture management: arthroscopically--assisted versus ORIF procedure – clinical and radiological comparison. Injury. 2017;48(Suppl 5):S61-S64.

2. Albuquerque RP, Hara R, Prado J, Schiavo L, Giordano V, Amaral NP. Estudo epidemiológico das fraturas do planalto tibial em Hospital de Trauma nível I. Acta Ortop Bras. 2013;21(2):109-15.

3. Prat-Fabregat S. Instructional lecture: knee treatment strategy for tibial plateau fractures: an update. EFORT Open Rev. 2016;1(5):225-32.

4. Elsoe R, Larsen P, Nielsen NPH, Swenne J, Rasmussen S, Ostgaard SE. Population-based epidemiology of tibial plateau fractures. Orthopedics. 2015;38(9):e780-6.

5. Mthethwa J, Chikate A. A review of the management of tibial plateau fractures. Musculoskelet Surg. 2017;102(2):119-27.

6. Molenaars RJ, Mellema JJ, Doornberg JN, Kloen P. Tibial plateau fracture characteristics: computed tomography mapping of lateral, medial, and bicondylar fractures. J Bone Joint Surg Am. 2015;97(18):1512-20.

7. Schatzker J. Anterior approach to the knee with osteotomy of the tibial tubercle for bicondylar tibial fractures. J Bone Joint Surg Am. 1988;70(2):208-19.

8. Schatzker J, McBroom R, Bruce D. The tibial plateau fracture. The Toronto experience 1968-1975. Clin Orthop Relat Res. 1979;(138):94-104.

9. 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.

10. Mandarino M, Pessoa A, Guimarães JAM. Avaliação da reprodutibilidade da classificação de Schatzker para as fraturas do Planalto Tibial. Rev INTO. 2004;2(2):11-8.

11. Luo CF, Sun H, Zhang B, Zeng BF. Three-column fixation for complex tibial plateau fractures. J Orthop Trauma. 2010;24(11):683-92.

12. Dirschl DR, Dawson PA. Injury severity assessment in tibial plateau fractures. Clin Orthop Relat Res. 2004;(423):85-92.

13. Martin J, Marsh JL, Nepola JV, Dirschl DR, Hurwitz S, DeCoster TA. Radiographic fracture assessments: which ones can we reliably make? J Orthop Trauma. 2000;14(6):379-85.