

## Research Article

# Inter- and Intraobserver Reliability for AO / ASIF and Moore Classification of Tibial Plateau Fractures – A Retrospective Study

Reiner Josef Wirbel<sup>1\*</sup>, Christina Vrabac<sup>2</sup>, Tim Pohlemann<sup>2</sup> and Sascha Hopp<sup>2</sup>

<sup>1</sup>Department of Trauma-, Hand- and Reconstructive Surgery, Germany

<sup>2</sup>Department of Trauma-, Hand- and Reconstructive Surgery, University Hospital of Saarland, Germany

**\*Corresponding author**

R. Wirbel, Department of Trauma-, Hand- and Reconstructive Surgery, Verbundkrankenhaus Bernkastel-Wittlich, Koblenzer Strasse 91, D-54516 Wittlich, Germany, Tel: 0049-6571-1532310; Fax: 0049-6571-1532390; E-mail: reiner.wirbel@web.de

Submitted: 03 May 2015

Accepted: 18 May 2015

Published: 19 May 2015

**Copyright**

© 2015 Wirbel et al.

**OPEN ACCESS****Keywords**

- Interobserver
- Intraobserver
- Reliability
- Tibial plateau fracture

**Abstract**

The AO / ASIF classification of tibial plateau fractures is based on radiographic morphological criteria, the classification according to Moore considers also functional criteria. The aim of this study was to compare the interobserver reliability and the intraobserver reproducibility of both classification systems.

Plain film radiographs and computed tomographs of 25 tibial plateau fractures were presented to 16 observers. There were three groups of observers with regard to their clinical expertise. Assessments were repeated 3 months later. The inter- and intraobserver reliability were evaluated by the kappa coefficients.

The interobserver reliability and the intraobserver reproducibility for both classification systems showed only fair to moderate agreement with kappa values ranged between 0.17 and 0.46. There were no considerable differences between the two classification systems. The experience of the observers did not influence the agreement at all. Even special training of the low experienced observers before the second assessment could not improve the interobserver agreement.

Reliability of the AO / ASIF and Moore classifications for tibial plateau fracture is challenging. Tibial plateau fractures are difficult to classify. Additional criteria have to be developed for reliable and reproducible classification. The results of clinical studies about tibial plateau fractures have to be analysed critically.

**ABBREVIATIONS**

AO: Arbeitsgemeinschaft für Osteosynthesefragen; ASIF: Association for the study of Internal Fixation

**INTRODUCTION**

Considering the severity of any fracture, every fracture classification has to provide a basis for decision making of treatment and for evaluation of the achieved results. It should be reliable and valid.

Tibial plateau fractures have a great variability and complexity. Three systems are commonly used for classification of tibial plateau fractures: the classification according AO / ASIF [1],

the Moore classification [2], and the Schatzker classification [3]. There exist only a few studies about the inter- and intraobserver reliability of the AO / ASIF and the Schatzker classification [3,4]. To our knowledge only two studies inform about the reliability of the Moore classification [5,6].

The interobserver agreement and the intraobserver reproducibility of different classification systems are determining for numerous fractures, such as fractures of the proximal femur, distal tibia, ankles, calcaneus, or proximal humerus [7-14].

Only fair to poor inter- and intraobserver reliability of the Neer and AO / ASIF classification for proximal humerus fractures is reported by [14]. Thus, the comparability of different clinical

studies is called into question especially according to the achieved results.

The aim of this study was to ascertain and compare the inter- and intraobserver reliability of the two classifications for tibial plateau fractures commonly used in our daily clinical practice: the AO / ASIF and the Moore classification. Furthermore, we wanted to check if the results were depended on the observer's level of professional experience.

## MATERIAL AND METHODS

### Classification systems

The AO / ASIF classification of tibial plateau fractures is based on radiographic morphological criteria [1]. Type-A-fractures comprise fractures in the extraarticular segment. Type-B-fractures are incomplete articular fractures, whereas type-C-fractures are complete articular fractures presenting a metaphyseal fracture line as well as an involvement of the articular surface. The details of the classification consisting of three fracture types and each with three groups are shown in Figure 1.

The classification according to Moore considers also clinical functional criteria, such as the injury mechanism [2]. "Fracture-dislocations" are distinguished from "stable" plateau fractures. "Fracture-dislocations" are generally unstable fractures and associated with a ligamentous injury. Thus, the preoperative appraisal of expected ligamentous, neuro-vascular or meniscal injuries of the knee joint will be possible. About 10-15% of all tibial plateau fractures are "fracture dislocations" and 85-90% is classified as stable plateau fractures. The latter fractures can present as cleavage fracture, pure depression fracture, combination of cleavage fracture and depression, or as bicondylar fracture. The bicondylar plateau fracture can be distinguished from the Moore fracture of type-5 by the absence of an additional eminentia fragment. The five types of "fracture-dislocations" according to Moore are summarized in Figure 2.

### Data records and observes

Ninety-one patients were treated for a tibial plateau fracture over a period of three years. The complete data sets of preoperative plain film radiographs and multidirectional, two-dimensional computed tomographs (CT) were available in 25 cases. The multidirectional CT-scans included two-dimensional axial, coronal, and sagittal views of the complete tibial plateau with slice thickness of 2 mm.

The data sets were admitted to 16 observers. There were three groups of observers with different professional experience. The observers of group I (n = 6) were senior consultants with a professional expertise in orthopaedic and trauma surgery of more than 15 years. We selected intentionally senior consultants from surrounding hospitals. All of them were heads of the department of traumatology and orthopaedic surgery. The observers of group II (n = 5) were senior registrars with a professional expertise of 6 to 10 years, whereas the observers of group III (n = 5) were fellows of the training program in trauma and orthopaedic surgery with a one to five year experience.

The data sets were admitted again to the observers after three

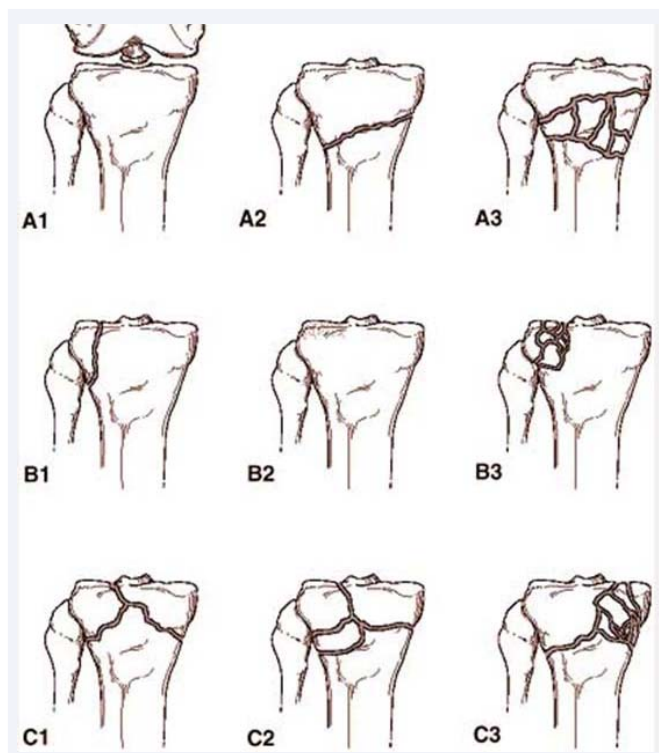


Figure 1 AO / ASIF classification of tibial plateau fractures.

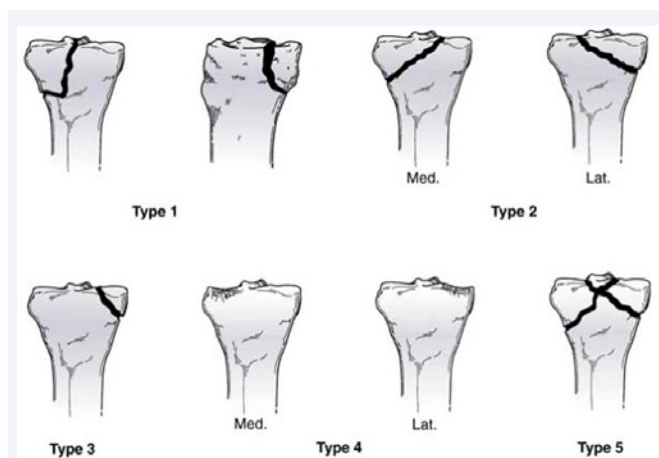


Figure 2 Moore classification of fracture-dislocations of the tibial plateau; Type I: coronal split fracture of the medial tibial plateau; Type II: entire condyle; Type III: rim avulsion; Type IV: rim compression; Type V: four-part fracture.

months. To eliminate the factor of recognition, the sequence of the data sets was modified. The observers should evaluate the tibial plateau fractures and classify according to the AO / ASIF criteria. There were nine options of classification concerning the categorization in one of the fracture type (A-C) and one of the groups (1-3). Additional classification according to Moore (type 1-5) should be performed, when a "fracture-dislocation" was assumed. The fellows of group III received a special training with instructions about the two classification systems for tibial plateau fractures after three months immediately before the second evaluation.

## Statistical analysis

Statistical analysis of the data obtained was performed using the software 2004 SAS 9.1.3 (SAS Institute Inc. Carry, NC, USA). For inter- and intraobserver reliabilities, the kappa statistic function were used measuring kappa values to describe the agreement between observers while correcting for the proportion that may have occurred by chance alone. A kappa value of 0 represented agreement by chance alone while kappa value of 1 implied a perfect agreement. Kappa values were interpreted using guidelines proposed by [15]. Values > 0.8 indicated excellent, 0.61 – 0.8 good, 0.41 – 0.6 moderate, 0.21 – 0.4 fair, and ≤ 0.2 poor reliability. The mean kappa values of every possible matched pair between and within the individual observer groups were obtained according to the method described by [16].

All of the 25 patients where the complete data sets of the plain films and CT-scans could be obtained, were informed, that their data would be submitted for publication, and gave their consent.

The study design was presented at the Institutional Ethical Board and the approval was given.

## RESULTS AND DISCUSSION

### Results

**Interobserver reliability:** The interobserver reliability of all observers revealed a kappa value of 0.29 and 0.34, respectively, at the two dates of data collection for the AO / ASIF classification. The kappa value was 0.23 and 0.31, respectively, for the Moore classification. In both classification systems, on average two fractures (8%) could not be classified by at least one of the observers. The interobserver kappa values within the different observer groups at the two assessment dates are shown in Table 1. We found only fair results. The values improved at the second assessment after three months, but the difference were not significant. The values were similar for both classification systems. It was noticeable, that the low experienced group

(group III) presented worse results for the Moore classification, but the difference was also not significant. The fact of a special training program for the fellows of group III immediately before the second assessment date provided only insignificantly better results for the AO / ASIF as well as for the Moore classification. Evaluation of the interobserver kappa values by comparing pairs of the observer groups at the two assessment dates showed that the highest agreement was between group I and group III (0.41) for the AO / ASIF classification. Altogether, the results have to be rated as fair to moderate presenting kappa values of 0.2 to 0.41 (Table 2).

**Intraobserver reliability:** The highest intraobserver kappa values were seen in group II (senior registrars) for the Moore classification. The lowest kappa value has to be observed in group III (fellows) for the AO / ASIF classification (Table 3).

### Discussion

The aim of every fracture classification is to offer helpful guidelines to the clinician. The communication about the fracture's severity, the prognostic value and the comparability of the achieved results should be provided. The two classification systems for tibial plateau fractures evaluated in the presented study were chosen because they are the ones that are mostly used in our clinical practice. By review of the literature there exist only two studies about the reliability of the Moore classification with different findings [5,6]. The kappa values for interobserver reliability range from 0.14 to 0.64 [5,6].

In our series, the kappa values of the inter- and intraobserver reliabilities showed only fair agreements for the AO / ASIF - as well as the Moore classification. Superiority of any of these two classification systems could not be seen.

It is well accepted that CT scan are better than plain film radiographs for analysing and classifying tibial plateau fractures [6,17]. It could demonstrated by [6] that three-dimensional CT scans in comparison with two-dimensional CT scans did not

**Table 1:** Interobserver kappa values within the different observer groups at the two assessment dates.

Group	Classification			
	AO / ASIF		Moore	
	T1	T2	T1	T2
I	0.31	0.40	0.17	0.35
II	0.23	0.19	0.27	0.26
III	0.30	0.36	0.19	0.25

Abbreviations: I: consultants; II: senior registrars; III: fellows; AO: Arbeitsgemeinschaft für Osteosynthesefragen; ASIF: Association for the study of Internal Fixation; T1: first assessment; T2: second assessment after 3 months

**Table 2:** Analysis of the interobserver kappa values for pairs at the two assessment dates.

Pairs / group	Classification			
	AO / ASIF		Moore	
	T1	T2	T1	T2
I versus II	0.29	0.33	0.27	0.33
I versus III	0.33	0.41	0.20	0.31
II versus III	0.26	0.30	0.25	0.31

Abbreviations: I: consultants; II: senior registrars; III: fellows; AO: Arbeitsgemeinschaft für Osteosynthesefragen; ASIF: Association for the study of Internal Fixation; T1: first assessment; T2: second assessment after 3 months

**Table 3:** Intraobserver kappa values between the two assessment dates.

Observer group	Classification	
	AO / ASIF	Moore
all	0.40	0.42
I	0.38	0.33
II	0.37	0.46
III	0.20	0.29

Abbreviations: I: consultants; II: senior registrars; III: fellows; AO: Arbeitsgemeinschaft für Osteosynthesefragen; ASIF: Association for the study of Internal Fixation

significantly improve the inter- and intraobserver reliability for characterization of tibial plateau fractures. Therefore, we chose for our study a complete set of two-dimensional CT scans including coronal, axial, and sagittal views. We did not put any time limit for appraisal of the complete data sets.

To ensure the external validity of the study and generalizability of the results we selected intentionally consultants from other hospitals representing one group of the observers.

There exist different statements in the literature about the impact of the professional expertise level on the inter- and intraobserver reliability for fracture classification. The significant impact of professional expertise on the intraobserver agreement is proved for certain fractures, such as distal radial fractures in childhood or proximal humeral fractures [12,13,18]. But the most studies state no significant impact of the professional experience on the inter- and intraobserver reliability, especially for classification of tibial plateau fractures [3,4,6,14,17,19,20]. This is in agreement with the findings in our study. Even special classification training of the low experienced observers before the second assessment could not significantly improve the intraobserver reliability. Further studies are intended to analyse if such an additional training of the experienced observers could maybe improve their intraobserver reliability.

The AO / ASIF – as well as the Moore classification for tibial plateau fractures showed only fair results for the inter- and intraobserver reliability. Is is reasonable to believe that the reliability of any fracture classification will improve in the clinical setting, where any information about the patient and the injury mechanism are available. This may be the reason for the minimal, but not significantly worse results for the Moore classification, which is based substantially on the injury mechanism.

Recent classification systems for tibial plateau fracture could not improve the reliability [4,21]. The classifications systems described by [4] and by [21] based on a three columns theory were not able to fulfil the expectations concerning good reliability values.

Reliability of classification of tibial plateau fractures is very difficult to achieve. Classification systems have to be as simple as possible to obtain a high interobserver reliability and intraobserver reproducibility [19,20]. But it seems to be obvious, that the complexity of tibial plateau fractures does not allow a simple categorization.

This has to be taken into consideration when analysing the results of clinical studies about tibial plateau fractures. The

comparability of clinical studies about tibial plateau fractures has to be questioned.

It is suggested, that new classification systems should combine the morphologic criteria of the AO-classification (i.e. extra-, partial intra- and completely intraarticular) with the clinically based criteria of the Moore-classification. The choice of the applied fixation method should also follow from the classification of tibial plateau fractures. For example, it seems to crucial in cases of intraarticular bicondylar fractures, if there is an additional central eminentia fragment, or the existence of a typical dorso-medial fragment, which requires an additional dorso-medial buttress plate. Thus, additional criteria have to be developed for reliable and reproducible classification of tibial plateau fractures.

## REFERENCES

- Müller ME, Nazarin S, Koch P, Schatzker J. The comprehensive classification of fractures of the long bones. Springer, Berlin, Heidelberg, New York, 1990.
- Moore TM. Fracture--dislocation of the knee. Clin Orthop Relat Res. 1981; 1: 128-140.
- 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: 396-398.
- Gicquel T, Najihi N, Vendevure T, Teyssedou S, Gayet LE, Hutten D. Tibial plateau fractures: reproducibility of three classifications (Schatzker, AO, Duparc) and a revised Duparc classification. Orthop Traumatol Surg Res. 2013; 99: 805-816.
- Maripuri SN, Rao P, Manoj-Thomas A, Mohanty K. The classification systems for tibial plateau fractures: how reliable are they? Injury. 2008; 39: 1216-1221.
- Doornberg JN, Rademakers MV, van den Bekerom MP, Kerkhoffs GM, Ahn J, Steller EP. Two-dimensional and three-dimensional computed tomography for the classification and characterisation of tibial plateau fractures. Injury. 2011; 42: 1416-1425.
- Schipper IB, Steyerberg EW, Castelein RM, van Vugt AB. Reliability of the AO/ASIF classification for perthrochanteric femoral fractures. Acta Orthop Scand. 2001; 72: 36-41.
- Ramappa M, Bajwa A, Singh A, Mackenney P, Hui A, Port A. Interobserver and intraobserver variations in tibial pilon fracture classification systems. Foot (Edinb). 2010; 20: 61-63.
- Swiontkowski MF, Sands AK, Agel J, Diab M, Schwappach JR, Kreder HJ. Interobserver variation in the AO/OTA fracture classification system for pilon fractures: is there a problem? J Orthop Trauma. 1997; 11: 467-470.
- Thomsen NO, Olsen LH, Nielsen ST. Kappa statistics in the assessment of observer variation: the significance of multiple observers classifying ankle fractures. J Orthop Sci. 2002; 7: 163-166.
- Humphrey CA, Dirschl DR, Ellis TJ. Interobserver reliability of a CT-based fracture classification system. J Orthop Trauma. 2005; 19: 616-622.
- Bahrs C, Schmal H, Lingenfelder E, Rolauffs B, Weise K, Dietz K. Inter- and intraobserver reliability of the MTM-classification for proximal humeral fractures: a prospective study. BMC Musculoskelet Disord. 2008; 9: 21.
- Brorson S, Bagger J, Sylvest A, Hróbjartsson A. Improved interobserver variation after training of doctors in the Neer system. A randomised trial. J Bone Joint Surg Br. 2002; 84: 950-954.

14. Siebenrock KA, Gerber C. The reproducibility of classification of fractures of the proximal end of the humerus. *J Bone Joint Surg Am.* 1993; 75: 1751-1755.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977; 33: 159-174.
16. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bulletin* 1971; 76: 378-382.
17. Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneeland JB, Dalinka MK. Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. *J Orthop Trauma.* 1997; 11: 484-489.
18. Randsborg PH, Sivertsen EA. Classification of distal radius fractures in children: good inter- and intraobserver reliability, which improves with clinical experience. *BMC Musculoskelet Disord.* 2012; 13: 6.
19. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand.* 2004; 75: 184-194.
20. Burstein AH. Fracture classification systems: do they work and are they useful? *J Bone Joint Surg Am.* 1993; 75: 1743-1744.
21. Luo CF, Sun H, Zhang B, Zeng BF. Three-column fixation for complex tibial plateau fractures. *J Orthop Trauma.* 2010; 24: 683-692.

#### Cite this article

Wirbel RJ, Vrabac C, Pohlemann T, Hopp S (2015) Inter- and Intraobserver Reliability for AO / ASIF and Moore Classification of Tibial Plateau Fractures – A Retrospective Study. *Ann Orthop Rheumatol* 3(2): 1048.