Inter- and intra-observer variation of the Schatzker and AO/OTA classifications of tibial plateau fractures and a proposal of a new classification system

CP CHARALAMBOUS, M TRYFONIDIS, F ALVI, M MORAN, C FANG, R SAMARAJI, P HIRST

Department of Orthopaedics and Trauma, Manchester Royal Infirmary, Manchester, UK

ABSTRACT

INTRODUCTION   The aim of this study was to evaluate the intra- and inter-observer variation of the Schatzker and AO/OTA classifications in assessing tibial plateau fractures, using plain radiographs.

PATIENTS AND METHODS   Fifty tibial plateau fractures were classified independently by six observers as per the Schatzker and AO/OTA classifications, using antero-posterior and lateral plain radiographs. Assessment was done on two occasions, 8 weeks apart.

RESULTS   We found that both the Schatzker and AO/OTA classifications have a high intra-observer (κ = 0.57 and 0.53, respectively), and inter-observer (κ = 0.41 and 0.43, respectively) variation. Classification of tibial plateau fractures into unicondylar versus bicondylar and pure splits versus articular depression ± split conferred improved inter- and intra-observer variation.

CONCLUSIONS   The high inter-observer variation found for the Schatzker and AO/OTA classifications must be taken into consideration when these are used as a guidance of treatment and when used in evaluating patients' outcome. Simply classifying tibial plateau fractures into unicondylar versus bicondylar and pure splits versus articular depression ± split may be more reliable.

CORRESPONDENCE TO

CP Charalambos, Department of Orthopaedics and Trauma, Manchester Royal Infirmary, Oxford Road, Manchester M13 9WL, UK
E: bcharalambos@hotmail.com

There is no universally accepted method of classification of tibial plateau fractures, with more than six classification schemes having been described. Of these, the Schatzker and AO/OTA classifications are the most commonly used methods for classifying such fractures.[1,2] There is little information regarding inter- and intra-observer variation when classifying tibial plateau fractures using the Schatzker and AO/OTA classification systems and hence this study was performed.

Patients and Methods

The Schatzker classification divides tibial plateau fractures into six types (Fig. 1). The AO/OTA classification divides proximal tibial fractures into types A, B and C. Each of the three types is divided into three groups described as 1–3, each of which having three further sub-groups. In this study, the broad AO/OTA classification consisting of the tibial plateau types and groups was used (Fig. 2). In the

AO/OTA classification, each group (*e.g.* B1, B2) is further subdivided into sub-groups(.1 to .3) but this division was not used for purposes of simplicity.

Fifty tibial plateau fractures presenting to our hospital over a 4-year period were used. All patients had anterior-posterior (AP) and lateral radiographs, as per hospital protocol. To ensure good quality radiographs, the hospital protocol requires the clinician assessing each patient to repeat any poor-quality radiograph. To determine intra- and inter-observer variation, each of six observers (two research fellows, two senior training orthopaedic surgeons [SpRs] and two lower limb orthopaedic and trauma consultants) independently assessed the AP and lateral radiographs of these 50 tibial plateau fractures and classified them according to the Schatzker and AO/OTA classifications. All participants in the study were familiar with both the Schatzker and AO/OTA classification systems. They were not given any clinical details regarding presentation or management of the
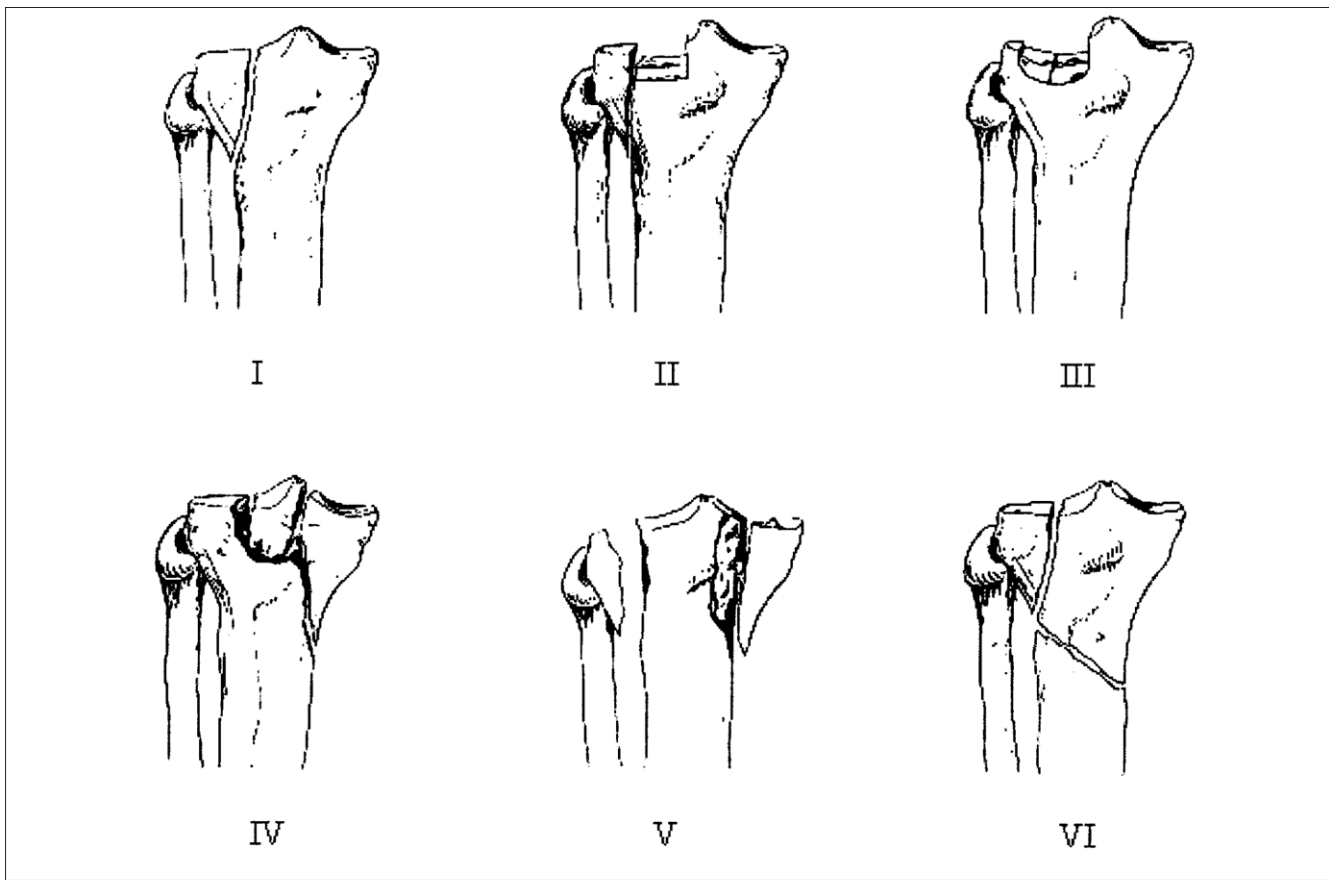
Figure 1 Schatzker classification for tibial plateau fractures.[2]

patients presenting with these fractures. Each observer was given a diagrammatic scheme and a written as well as verbal description of the Schatzker and AO/OTA classifications. They were given as much time as they required to evaluate the radiographs accurately. The observers indicated their choices on a pre-designed proforma having schematic representation of the Schatzker and AO/OTA classification. The series was arranged randomly, numbered 1 to 50, and included all different patterns of tibial plateau fractures. All radiographs were anonymous. All observers evaluated the radiographs on two occasions, 8 weeks

apart. The classification choices made at the first viewing were not available during the second viewing. The observers were not provided with any feedback after the first viewing and the radiographs were not available to any of them between the first and second viewings. Ethical approval was not required by our institution for this type of study, at the time this was performed.

For statistical analysis the κ-test of Cohen[3] was used to determine the level of variation. Kappa is a coefficient of agreement, that varies from +1 (perfect agreement) to '0' (agreement no better than chance), to –1 (representing absolute
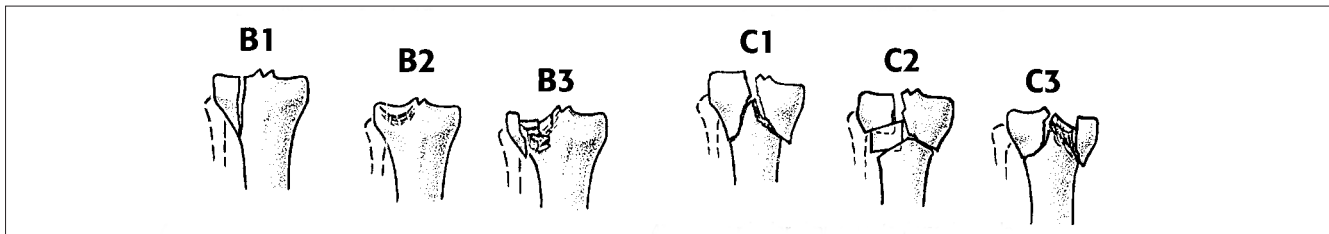

Figure 2 Broad AO/OTA classification for tibial plateau fractures: B1, split fractures only; B2, depressed fractures only; B3, combined split and depression; C1, articular simple, metaphyseal simple; C2, articular simple, metaphyseal multifragmentary; and C3, multifragmentary articular.

Table 1   Kappa values for inter- and intra-observer variation for the Schatzker classification

	Inter-observer						Intra-observer
	A	B	C	D	E	F	
A		0.42	0.50	0.29	0.39	0.45	0.53
B			0.53	0.40	0.32	0.47	0.65
C				0.46	0.38	0.55	0.70
D					0.32	0.33	0.36
E						0.38	0.51
F							0.66
Mean κ value for each observer	0.41	0.43	0.48	0.36	0.36	0.44	
Mean κ value for all observers			0.41				0.57

A and B were consultants, C and D were SpRs, E and F were research fellows.

disagreement). The results of the first reading were used to determine inter-observer variation. Comparison of the first and second readings was determined intra-observer variation.

Inter- and intra-observer variations were determined initially for the Schatzker and AO/OTA classification. Using the responses given to Schatzker and AO/OTA classification, we looked at the inter- and intra-observer variation in classifying tibial plateau fractures as unicondylar versus bicondylar and pure split versus articular surface depression ± split. For distinguishing unicondylar versus bicondylar fractures we determined for the AO/OTA classification the ability to distinguish B1, B2, B3 fractures from C1, C2, C3 and for the Schatzker I, II, III, IV fractures from V and VI. For distinguishing pure split versus articular surface depression ± split, we used for the AO/OTA classification B1 versus B2, B3 and for the Schatzker I versus II, III.

Table 2   Kappa values for inter- and intra-observer variation for the AO classification

	Inter-observer						Intra-observer
	A	B	C	D	E	F	
A		0.56	0.58	0.24	0.51	0.47	0.51
B			0.50	0.34	0.47	0.43	0.61
C				0.34	0.41	0.45	0.60
D					0.35	0.36	0.37
E						0.43	0.54
F							0.58
Mean κ value for each observer	0.47	0.46	0.46	0.33	0.43	0.43	
Mean κ value for all observers			0.43				0.53

A and B were consultants, C and D were SpRs, E and F were research fellows.

Results

Tables 1 and 2 give the results of the κ statistical analysis of inter- and intra-observer variation for the Schatzker and AO/OTA classifications, respectively. The range of values for inter-observer variation using the Schatzker classification was from 0.29 to 0.55, with an overall mean of 0.41 (SD 0.08). Intra-observer analysis for the Schatzker classification gave κ values ranging from 0.36 to 0.7 with a mean of 0.57 (SD 0.13). The inter-observer variability for the AO/OTA classification system yielded κ values from 0.24 to 0.58 with a mean of 0.43 (SD 0.09). Intra-observer analysis gave κ values ranging from 0.37 to 0.61 with a mean of 0.53 (SD 0.09). There was no statistically significant difference in the inter- and intra-observer variation between the Schatzker and AO/OTA classification systems ($P = 0.47$ and $P = 0.57$, respectively; Mann-Whitney test). In view of the above results, we then looked at the inter- and intra-observer variation in classifying tibial plateau fractures as unicondylar versus bicondylar and pure split versus articular surface depression ± split. The inter- and intra-observer κ values in distinguishing unicondylar versus bicondylar fractures were for the AO/OTA classification (B1, B2, B3 versus C1, C2, C3) 0.62 and 0.77, respectively, and for the Schatzker (I, II, III, IV versus V, VI) 0.67 and 0.70, respectively. The inter- and intra-observer κ values in distinguishing pure split versus articular surface depression ± split was for the AO/OTA classification (B1 versus B2, B3) 0.51 and 0.61, respectively, and for the Schatzker (I versus II, III) 0.60 and 0.62, respectively.

Discussion

Intra- and inter-observer variation is essential for any classification system. We found the correlation (κ) values for both the intra- and inter-observer variation to be low when using either the Schatzker or the AO/OTA classification systems for tibial plateau fractures. There have been previous attempts to qualify variation depending on the κ values. Svansholm *et al.*[4] arbitrarily divided variation as poor (κ < 0.5), good (κ 0.5–0.75) and excellent (κ > 0.75). We feel that such division is arbitrary and can give the wrong message when the κ values are borderline; hence, we did no use it.

In principle, both the Schatzker and the broad AO/OTA classifications for tibial plateau fractures are simple, as they only have six divisions each. The high inter- and intra-observer variation in this study may be related to the fact that tibial plateau fractures are often complex injuries with multiple fracture lines and variable articular line depression, which may be difficult to assess radiologically. CT or MRI scans may help to improve the sensitivity and specificity of identifying the extent of these fractures, and might help to improve inter- and intra-observer classification variation.[5–7]

Table 3 Proposed new classification for tibial plateau fractures. Fractures are described according to the number of condyles involved as well as the split/articular surface depression combination

Number of condyles involved
Unicondylar (medial or lateral)
Bicondylar (medial or lateral)

Type of fracture
Pure split
Articular surface depression without split
Articular surface depression ± split

However, their use is not widely considered part of the standard evaluation of these fractures; in clinical practice, plain radiographs are often used in isolation to decide upon treatment. For cases where surgery is performed, operative findings may further help classify tibial plateau fractures. In cases where non-operative treatment is decided, plain radiographs may be the only means for classification. It is for these reasons that we used plain radiographs rather than further imaging for fracture classification. It should be noted that both the Schatzker and AO classifications were originally based on plain radiographs, although more recently CT scanning has also been used in classifying tibial plateau fractures. Although, we used both AP and lateral radiograph trauma series, the inter-observer variation coefficient did not reach the high levels that would be expected for such widely used and accepted classification schemes. In the current study, intra-observer reproducibility was found to be higher than inter-observer variation, which is similar to that reported for other classification systems. This is because intra-observer reproducibility reflects reproducibility independent of agreement. As a result, incorrect responses can still give good intra-observer reproducibility even though this may reflect that the observer is consistently wrong. Thus incorrect responses may show lower intra-observer than inter-observer variation.

The level of expertise did not seem to be an important factor. Of the six observers, one of the SpRs obtained the highest correlation coefficient (κ 0.7) for the Schatzker, and a consultant (κ 0.61) for the AO/OTA classification system. The lowest correlation coefficient was obtained by an SpR for the Schatzker (κ 0.36) and AO/OTA (κ 0.37) classification.

In managing tibial plateau fractures, important fracture features are whether they are unicondylar or bicondylar and whether they involve a split fracture or pure articular

depression. Differentiating between bicondylar and mono-condylar fractures is important because it can guide the type of surgical fixation (ring fixator, or double buttress plating for bicondylar versus screw fixation or buttress plating for uni-condylar). Differentiating between pure split fractures and those consisting of articular depression with or without split is important as surgery for the latter needs to involve reduction and restoration of the articular surface. We have shown that simply classifying fractures as unicondylar or bicondylar and further dividing the unicondylar fractures as pure split or articular depression with or without split improves both the inter- and intra-observer variability. On the basis of this we propose a simple descriptive classification for tibial plateau fractures (Table 3). We feel that the new classification system proposed can help to guide management. However, whether such a system will help to predict prognosis can only be examined by a prospective clinical trial.

Conclusions

Our findings suggest a high intra- and inter-observer variation in classifying tibial plateau fractures using the Schatzker or broad AO/OTA systems with plain radiographs. Classification into unicondylar versus bicondylar and pure splits versus articular depression ± split may confer improved inter- and intra-observer agreement.

References

1. Müller ME, Nazarian S, Koch P, Schatzker J. *The comprehensive classification of fractures of long bones.* New York: Springer, 1990, 148–56.

2. Schatzker J, McBroom R, Bruce D. Tibial plateau fracture. *Clin Orthop* 1979; **138**: 94–104.

3. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Measure* 1960; **20**: 27–46.

4. Svanholm H, Starklint H, Gundersen HJG. Reproducibility of histomorphologic diagnoses with special reference to the κ statistic. *APMIS* 1989; **97**: 689–98.

5. Holt MD, Williams LA, Dent CM. MRI in the assessment of tibial plateau fractures. *Injury* 1995; **26**: 595–9.

6. Chan PS, Klimkiewicz JJ, Luchetti WT, Esterhai JL, Kneelend JB, Heppenstal RB. Impact of CT scan on treatment plan and fracture classification of tibial plateau fractures. *J Orthop Trauma* 1997; **11**: 484–9.

7. Martin J, Marsh JL, Nepola JV, Dirschl DR, Hurwitz S, De Coster T. Radiographic fracture assessments: which ones can we reliably make? *J Orthop Trauma* 2002; **16**: 632–7.