Orthopaedics & Traumatology: Surgery & Research (2013) 99, 805—816

Available online at
SciVerse ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

ORIGINAL ARTICLE

Tibial plateau fractures: Reproducibility of three classifications (Schatzker, AO, Duparc) and a revised Duparc classification

T. Gicquel[a], N. Najihi[a], T. Vendeuvre[b], S. Teyssedou[b], L.-E. Gayet[b], D. Huten[a],*

[a] *Service d'orthopédie-traumatologie, hôpital Pontchaillou, 2, rue Le Guilloux, 35033 Rennes, France*
[b] *Service d'orthopédie-traumatologie, CHU de Poitiers, 86000 Poitiers, France*

Accepted: 7 June 2013

KEYWORDS
Tibial plateau fractures;
Classification system;
Reproducibility;
CT scan

Summary
Introduction: Since the reproducibility of the Schatzker and AO tibial plateau fracture classification systems has already been assessed, the goal of this study was to evaluate the Duparc classification system and compare it to the other two.
Hypotheses: CT scan is better than X-rays for analyzing and classifying tibial plateau fractures. The Duparc classification system is more effective than the other two systems but could be improved by adding elements of each.
Materials and methods: Six observers analyzed images from 50 fractures and then classified them. Each fracture was evaluated on X-rays. Two weeks later, these same fractures were evaluated on X-rays and CT scans. The same process was repeated four weeks later. The Kappa coefficient (κ) was used to measure agreement and contingency tables were built.
Results: The interobserver reproducibility for the X-ray analysis was poor for the Duparc and AO classifications ($\kappa_{Duparc} = 0.365$; $\kappa_{AO} = 0.357$) and average for the Schatzker classification ($\kappa_{Schatzker} = 0.404$). The reproducibility was improved overall when CT scans were also analyzed ($\kappa_{Duparc} = 0.474$; $\kappa_{AO} = 0.479$; $\kappa_{Schatzker} = 0.476$). A significantly greater number of fractures could not be classified in the Schatzker system than in the others (14.3% versus 2% for Duparc and 7.33% for AO). Review of the contingency tables revealed that the Schatzker and AO classification systems did not take certain fracture types into account. Seventy-one percent (71%) of the lateral unicondylar split fractures were found to be combined fractures when CT scan analysis was added.
Discussion: Our results showed CT scan to be better at analyzing and classifying fractures. We also found the Duparc classification to be advantageous because it allowed more fractures to

* Corresponding author. Tel.: +33 02 99 28 43 21.
E-mail address: denis.huten@chu-rennes.fr (D. Huten).

1877-0568/$ – see front matter
http://dx.doi.org/10.1016/j.otsr.2013.06.007

be classified than in other classification systems, while having similar reproducibility. Based on our study findings, the Duparc classification was revised by adding elements of the other two. We propose using the modified Duparc classification system to analyze tibial plateau fractures going forward.

Level of evidence: Level IV. Retrospective study.

Introduction

Tibial plateau fractures must be properly identified before they can be treated. The first classification system was proposed by Marchant [1], who described three fracture types: split, depression and combined. The 1960 Duparc and Ficat classification [2] (revised in 1990 [3]) is used in France. The Schatzker classification system [4] is the most commonly used in English-language and international publications. The AO classification system [5] is one part of a general alphanumeric classification system for all fractures. Other existing classification systems are not widely used [6,7]. The performance of the Schatzker and AO systems has already been studied [8—15]. Results vary depending on the imaging modality used; CT scan has been shown to improve reproducibility. However, the Duparc classification has not been evaluated to the same degree.

The main goal of this study was to compare these three classification systems by evaluating their intra- and interobserver reproducibility with conventional X-rays then with CT scan and then determining their ability to classify as many fractures as possible, to determine which system is the most relevant. We hypothesized that CT scans would be better than conventional X-rays and that the Duparc classification would be the most relevant.

Material and methods

Only recent tibial plateau fractures in adults having good quality X-rays and CT scans were included. Intercondylar eminence and tibial tuberosity fractures were excluded. Of the 117 records from various hospital centers in France (Angers, Caen, Nantes, Poitiers, Tours, Rennes) meeting these criteria, 50 were randomly selected in accordance with similar published studies [8—15].

Two digital imaging files were created for each fracture. One file contained the AP and lateral X-rays (¾ views were not always available) and was called the ''X-ray'' file (XR). The other file contained the same X-rays plus six axial, six coronal and six sagittal CT slices and was called the ''X-ray with CT'' file (XR/CT). All files were made anonymous and randomly numbered within the two groups (XR and XR/CT) so that no pattern was apparent.

Six observers from Rennes and Poitiers (1 university professor/staff physician, 1 fellow and 1 resident at each center) analyzed and then classified each fracture. None had been involved in treating these fractures or in selecting the images. The data was collected in an Excel spreadsheet with drop-down lists for each response. To standardize the answers, a user manual was given to each observer with reminders of the classification systems (diagrams and written descriptions) and detailed information on the study-related items and potential answers.

Injury features were described with 22 items (Table 1). The Duparc classification (Fig. 1) consisted of five fracture types (lateral unicondylar, medial unicondylar, bicondylar, spinocondylar, posteromedial) and 16 sub-types; the Schatzker classification (Fig. 2) had six types, and the AO classification (Fig. 3) had 7 types (A was excluded; B1, B2, B3, C1, C2, C3 were included) and 14 sub-types. Each fracture was classified (or not classified) among the types and sub-types in the Duparc and AO systems and the types in the Schatzker system by the six observers.

Each observer analyzed the XR file and then the XR/CT file two weeks later (first round) to evaluate the relative contribution of CT scanning. The entire process was repeated four weeks later (second round). Each analysis comprised 300 answers. The interobserver reproducibility was calculated on the data from the first round to avoid recall bias. The averages of all intra- and interobserver Kappa coefficients were calculated and compared using Student's *t*-test (paired when appropriate).

The Kappa was calculated by taking into consideration the Duparc and AO sub-types and the Schatzker types, and then the Duparc types (simplified Duparc classification) and AO types (simplified AO classification) to have the same or nearly the same number of responses for each classification system. The Kappa coefficient [16] reflects how many responses the observers agreed on and how many agreements occurred by chance [17]. When there is 100% agreement, it has a value of 1.00 (maximum); when the agreement is attributed only to chance, its value is 0 (minimum). The values were interpreted according to Landis and Koch [18]: <0.21 slight; 0.21—0.40 fair; 0.41—0.60 moderate; 0.61—0.80 substantial; 0.81—1.00 excellent.

Contingency tables (cross tabulations) were built using the 300 XR/CT evaluations in the first round and the 137 fractures that were classified as lateral unicondylar during the first round. The rate of non-classified fractures for each classification system was determined from the sum of ''non-classified'' responses during the first round and statistically evaluated with a Z-test. The statistical analysis was performed with XL Stat software (Addinsoft©, New-York, NY, USA). Significance threshold was set at 0.01.

Results

Duparc classification

The interobserver correlation was fair with XR ($\kappa_{XR} = 0.365$) and moderate with XR/CT ($\kappa_{XR/CT} = 0.474$) (Table 2). The