ORIGINAL ARTICLE

CT scan does not improve the reproducibility of trochanteric fracture classification: A prospective observational study of 53 cases

E. Cavaignac\textsuperscript{a,}\textsuperscript{*}, M. Lecoq\textsuperscript{a}, A. Ponsot\textsuperscript{b}, A. Moine\textsuperscript{a}, N. Bonnevialle\textsuperscript{a}, P. Mansat\textsuperscript{a}, N. Sans\textsuperscript{b}, P. Bonnevialle\textsuperscript{a}

\textsuperscript{a} Musculoskeletal Institute at Toulouse, Teaching Hospital Center, Purpan Orthopaedics and Trauma Unit, place Baylac, 31052 Toulouse, France
\textsuperscript{b} Central Radiology Department at Purpan, Teaching Hospital Center, place Baylac, 31052 Toulouse, France

Accepted: 21 September 2012

KEYWORDS
Trochanteric fractures; CT scan; 3D reconstruction

Summary

Introduction: The reproducibility of various classification systems for trochanteric fractures is poor. This problem could be related to a lack of readability when using conventional X-rays.

Hypothesis: Computed tomography scanning (CT scan) will improve the interobserver reproducibility of the AO classification for trochanteric fractures.

Patients and methods: This was a prospective, observational, descriptive study following a group of 53 patients with trochanteric fractures. The fractures were evaluated with conventional X-rays, CT scan slices and 3D reconstruction (3DR). The resulting images were blinded and analysed by two observers using two classification systems: AO and Evans modified by Jensen (EVJE). A sample size of 53 was needed to show an improvement in the interobserver reproducibility when deciding the AO classification type with CT scan images. Kappa coefficients were used to measure interobserver reproducibility and agreement; agreement is the degree of consistency in the analysis by one observer who views the same fracture on two different imaging modalities.

Results: The interobserver reproducibility for the AO classification was 0.28 for X-rays, 0.33 for CT scan and 0.28 for 3DR. For the EVJE classification, these coefficients were 0.50 for X-rays, 0.35 for CT scan and 0.47 for 3DR. The agreement rate between the two imaging modalities was between 0.38 and 0.58 for X-rays/CT scan and between 0.79 and 0.86 for CT scan/3DR.

Discussion: The primary objective of this study was not achieved. CT imaging does not improve the interobserver reproducibility of various classification systems for trochanteric fractures.

* Corresponding author.
\textit{E-mail address: cavaignac.etienne@gmail.com} (E. Cavaignac).
Introduction

Multiple classification systems, most of them descriptive [1–6], have been put forward to differentiate between various trochanter area fractures, which have become more common as the population ages [7]. The most used classification systems are the one by Evans [5], modified by Jensen et al. [2] and the one by Muller et al. [1] (AO). Despite routine use, the interobserver reproducibility of these two systems is poor [8–12]. Also, the classification is based on single plane, A/P views from X-rays. The true trajectory of the fracture line (especially in the sagittal plane) and the degree of comminution are difficult to evaluate. But this information is important due to the potential impact on the mechanical stability of these injuries, which are mainly treated by internal fixation. Parker and Handoll [13] have shown that intramedullary devices are more effective than extramedullary devices, especially for fractures involving the lateral cortex of the greater trochanter. Gotfried [14] and then Palm et al., [15] have pointed out that extramedullary fixation is only effective when the lateral trochanteric cortex is intact.

The hypothesis driving the current work was that computed tomography scanning (CT scan), because of its more extensive analysis of bone anatomy, will improve the reproducibility rates for the Evans-Jensen (EVJE) and Muller (AO) classification systems.

The primary goal of this study was to show an improvement in the interobserver reproducibility when CT scan images are used to establish the fracture type in these two classification systems. The secondary goals of this study were to analyse the ability of conventional X-rays to detect lateral cortex involvement and to measure the reproducibility of the two above-mentioned classification systems.

Patients and methods

Patients

We performed a prospective, single-centre, observational, descriptive study of patients presenting with a trochanteric fracture. This protocol was not submitted for approval at our facility. Between January 1, 2011 and April 30, 2011, patients who presented at our emergency ward with a recent trochanter area fracture were proposed for inclusion into an imaging protocol, as long as a CT scanner and a radiologist (AP and NS) were available. Patients with fractures in pathological bone (metastasis, bone pathology) were excluded. The trochanter area was defined, based on AO classification criteria, as the area located between the base of the femoral neck in the cranial orientation and a horizontal line passing 2.5 cm below the lesser trochanter in the caudal orientation, which forms the “epiphyseal square” [1]. The imaging assessment consisted of A/P X-rays of the injured hip obtained in all cases, lateral X-rays obtained in 30 cases (56%), CT scanning with axial slices (CT scan) and with three-dimensional reconstruction (3DR). The CT scanning was performed with a four-slice CT scanner (Siemens AG, Erlangen, Germany). The 3D reconstructions were performed by a radiologist (AP) from contiguous slices at 2 mm intervals, providing a 360º view based on 36 images.

During this period, 58 patients were eligible for the study. Five were excluded from this study because the paraclinical assessment was incomplete: in four cases, the physician at admission was not aware of the study protocol and, in one case, the CT scanner was not available in a reasonable time frame before the surgery. Thus, 53 patients (41 women, 12 men) with an average age of 84 years (SD: 10.8) were included. The fracture was on left side in 26 cases and on the right side in 27 cases.

Methods

For each patient, the images obtained were made anonymous and analysed independently. The two observers (ML and EC) were asked to classify each fracture based on the AO and EVJE classification systems with three different imaging modalities (X-rays, CT scan, 3DR), interpreted in isolation and without knowledge of the results from the other images. The AO classification has three main groups that are subdivided into three subgroups [1]. The EVJE classification consists of five pertochanteric fracture types and one R type (for reverse) that groups all intertrochanteric fractures [2, 5]. The analysis was performed on a series of 20 continuous images (X-rays, CT scan, 3DR) that was repeated until the full series had been analysed. Any fracture in the lateral cortex of the greater trochanter was noted. This area, called the “lateral wall” by Gotfried [14], corresponds to the lateral trochanteric cortex going from the top of the greater trochanter to a horizontal line that crosses the lower margin of the lesser trochanter.

Statistical methods

The various items were recorded in spreadsheet software (Excel, Microsoft, Redmond, WA, USA). The descriptive analysis revolved around distributing the population into the various groups by the two observers based on the various imaging modalities, along with the presence or absence of a lateral wall fracture. The comparative analysis revolved around the measurement of interobserver reproducibility and agreement in the classification between the various
imaging modalities using Cohen's kappa coefficient. Reproducibility and agreement were considered better as the coefficient approached a value of 1. They were deemed excellent when between 1 and 0.81; good when between 0.80 and 0.61; average when between 0.60 and 0.41; weak or poor when below 0.40 (based on Landis and Koch [16]). We considered CT scanning as the gold standard for bone imaging and then determined the diagnostic ability of conventional X-rays (sensitivity, specificity, positive and negative predictive value) in detecting lateral wall involvement. The sample size calculation was made to be able to demonstrate an interobserver reproducibility of 0.61 (good agreement) for the CT scan and the AO classification system with a Type I error risk of 5% and a power of 80%. This led to 53 subjects being included in this study. The calculation was based on the study by Pervez et al., [11] (5 observers, 88 subjects) who found an interobserver agreement kappa of 0.42 for an analysis of the AO classification.

Results

Based on conventional X-rays and using the AO classification (Fig. 1), ML put 14 fractures (27%) in group A1, 31 (57%) in group A2 and eight (15%) in group A3. With the same classification, EC put 15 fractures (28%) in group A1, 30 (58%) in group A2 and eight (15%) in group A3. Based on the analysis of axial CT slices (Fig. 2), ML put eight fractures (13%) in group A1 of the AO classification, 38 (72%) in group A2 and eight (15%) in group A3; for EC the distribution was eight in A1 (15%), 38 in A2 (72%) and nine in A3 (17%). Based on 3D reconstruction images (Fig. 3), ML put nine fractures (16%)

Figure 1 Results of the analysis using conventional X-ray images according to the AO classification by two observers for 53 trochanteric fractures. Black bars correspond to EC (Junior) and grey bars correspond to ML (Senior). The bar graph shows the population distribution in the nine AO subgroups; the table provides the population distribution in percentages in the A1, A2 and A3 groups for the two observers.

Figure 2 Results of the analysis using transverse axial CT slice (CT scan) images according to the AO classification by two observers for 53 trochanteric fractures. Black bars correspond to EC (Junior) and grey bars correspond to ML (Senior). The bar graph shows the population distribution in the nine AO subgroups; the table provides the population distribution in percentages in the A1, A2 and A3 groups for the two observers.

Figure 3 Results of the analysis using 3D reconstruction (3DR) images according to the AO classification by two observers for 53 trochanteric fractures. Black bars correspond to EC (Junior) and grey bars correspond to ML (Senior). The bar graph shows the population distribution in the nine AO subgroups; the table provides the population distribution in percentages in the A1, A2 and A3 groups for the two observers.
Based on conventional X-rays (Fig. 4A) and using the EVJE classification, ML classified four fractures (8%) as Type 1, 10 as Type 2 (19%), five as Type 3 (9%), nine as Type 4 (17%), twenty as Type 5 (38%), and five as Type R (9%). For EC, the distribution was six as Type 1 (11%), eight as Type 2 (15%), eight as Type 3 (15%), four as Type 4 (8%), twenty as Type 5 (38%) and seven as Type R (13%). Based on the analysis of axial CT slices (Fig. 4B), ML classified three fractures as Type 1 (5%), four as Type 2 (8%), eight as Type 3 (15%), seven as Type 4 (13%), 25 as Type 5 (51%) and four as Type R (8%). For EC the distribution was four as Type 1 (8%), five as Type 2 (9%), six as Type 3 (11%), seven as Type 4 (13%), 22 as Type 5 (42%) and nine as Type R (17%). Based on 3D reconstruction images (Fig. 4C), ML classified three fractures as Type 1 (5%), six as Type 2 (11%), eight as Type 3 (15%), eight as Type 4 (15%), 24 as Type 5 (46%), and four as Type R (8%). For EC, the distribution was (Fig. 4C) four as Type 1 (8%), three as Type 2 (5%), nine as Type 3 (17%), seven as Type 4 (13%), 21 as Type 5 (40%) and nine as Type R (17%).

The lateral wall was determined to be fractured on 10 conventional X-rays by the two observers. With CT scan images, EC found 16 lateral wall fractures and ML only 13. In every case, the lateral wall fracture occurred in the context of an EVJE Type 5 or Type R fracture.

The Kappa coefficients for the agreement between the various imaging modalities for ML and EC with the AO classification system are summarized in Table 1. For the comparison between conventional X-rays (X-rays) and axial CT slices (CT scan), the agreement between the two modalities was average for ML (kappa = 0.58) and poor for EC (kappa = 0.38). For the comparison between CT scan and 3DR, the agreement was excellent for ML (kappa = 0.85) and good for EC (kappa = 0.79). The Kappa coefficients for the agreement between the various imaging modalities for ML and EC with the EVJE classification systems are summarized in Table 2. For the comparison between X-rays and CT scan, the agreement was average for the two observers (kappa of 0.56 for ML and 0.58 for EC). For the comparison between CT scan and 3DR, the agreement was excellent for ML (kappa = 0.86) and good for EC (kappa = 0.80).

The interobserver kappa coefficients for the two classification systems are summarized in Table 3. When using the AO classification and conventional X-rays, axial CT slices or 3D reconstruction, the interobserver correlation was consistently poor (kappa of 0.28, 0.33 and 0.28, respectively). When using the EVJE classification system, the interobserver correlation was average when based on conventional X-rays (kappa = 0.50), poor when based on axial

### Table 1
<table>
<thead>
<tr>
<th>Observer</th>
<th>X-rays/CT scan</th>
<th>CT scan/R3D</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML (senior)</td>
<td>0.58</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>0.46–0.70</td>
<td>0.73–0.99</td>
</tr>
<tr>
<td>EC (junior)</td>
<td>0.38</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>0.27–0.48</td>
<td>0.68–0.91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Classification</th>
<th>Conventional X-rays</th>
<th>CT scan slices</th>
<th>3D Reconstruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>AO</td>
<td>0.28</td>
<td>0.33</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>0.17–0.39</td>
<td>0.21–0.45</td>
<td>0.17–0.40</td>
</tr>
<tr>
<td>EC</td>
<td>0.5</td>
<td>0.35</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>0.37–0.62</td>
<td>0.21–0.49</td>
<td>0.34–0.60</td>
</tr>
</tbody>
</table>

EVJE: Evans modified by Jensen; CT scan: computed tomography scanning.
CT slices (kappa = 0.35) and average when based on 3DR (kappa = 0.47).

For the detection of lateral wall fracture on X-rays, the analysis by ML resulted in a sensitivity of 77%, specificity of 100%, positive predictive value of 100% and negative predictive value of 93%. For EC, the various results were 62.5%, 100%, 100% and 86%, respectively. The interobserver reproducibility for the assessment of lateral wall fracture was excellent with conventional X-rays (kappa = 1.0), axial CT slices (kappa = 0.95) and 3D reconstructions (kappa = 0.95). The agreement was also very high, with a kappa coefficient of 0.90 for EC and 0.95 for ML between the conventional X-rays and axial CT slices; the kappa coefficient was 1.0 for both observers between the axial CT slices and 3D reconstruction.

Discussion

The results of this study did not support our hypothesis and the primary goal of this study was not attained — the kappa coefficient for interobserver reproducibility, for the AO classification system, did not increase to the degree expected (from 0.28 to 0.33). Based on the study by Pervez et al., [11], we had expected to reach a coefficient of 0.61, which was not achieved when adding assessments based on axial CT slices or 3D reconstruction. The current series had enough patients to demonstrate that a significant increase in the interobserver reproducibility could be achieved by using an imaging modality with better anatomical definition. The use of CT scanning to look for complex fractures at other sites is standard and even systematic [17–22]. Thus, we did not deem it necessary to get approval from the research ethics committee at our institution to give patients this extra dose of radiation.

This study has several limitations that cannot be ignored. Interpretation of conventional X-rays is performed on images taken in an emergency setting, thus the quality of these images may leave something to be desired. Koval et al., [23] proposed performing the X-rays with traction and internal rotation of the leg to improve the precision of the diagnosis and especially to reduce the errors in distinguishing between femoral neck and trochanteric fractures; this was not a problem in the current study. Three-dimensional reconstructions can have an artificial appearance, as certain parts of the image become smooth, and some small comminuted cortical fragments can be seen. We decided that only two observers were needed for all the X-rays. The goal of the study was not to establish the reproducibility rate of the various classification systems, but to determine if the reproducibility could be improved by using a more precise imaging modality. The reproducibility results in the current study are consistent with those reported in the published literature with more than two observers [8–11, 24]. Also, one of the observers, being a junior surgeon, does not limit the quality of the study, as explained by Van Embden et al., [8]. The process followed in this study, where fracture classification established on conventional X-rays is extended to CT scan analysis, has been done with fractures in other locations [17–22].

One of the novel findings of the current study was that fracture comminution is common and under-estimated by conventional X-rays. This new result was acquired with the axial CT slices; 3D reconstructions did not provide additional information but were visually pleasing. Our protocol did not improve the reproducibility of the EVJE classification on axial CT slices — it actually reduced it. These two classification systems do not have a readability problem, but one of understanding and analysis, as they were established, based on a single image and rendered as planar drawings [1, 2, 5].

The current distribution of trochanteric fractures corresponds to previously published reports [25]. Eight out of 10 (80%) are pertrochanteric, from the top of the greater trochanter to the lesser trochanter, and often comminuted. Intertrochanteric fractures, having a horizontal fracture line from the lateral cortex to the lesser trochanter, make up about 15% of trochanteric fractures. The latter fractures are at first sight, easy to identify in the two classification systems (A3 and R), however the two observers were not always in agreement as to the appropriate AO classification when using conventional X-rays (Fig. 1). The interobserver reproducibility for the EVJE classification was higher than the one obtained for the AO classification. For the EVJE classification, the complexity of the intertrochanteric fracture line is not taken into account like it is in the AO classification; significant comminution or multiple fracture lines makes it difficult to classify the fracture as pertrochanteric or intertrochanteric. The two observers were unanimous in identifying fracture comminution on CT scan images more often than on conventional X-ray images for both classification systems. Similarly, the two observers changed the fracture identification minimally when going from axial CT slices to 3D reconstructions for both classification systems.

In this study, the interobserver reproducibility and agreement were limited for the two classification systems. These results were consistent with those reported in the published literature. When using the Evans classification, Andersen et al., [24] found an interobserver kappa of 0.38 and intra-observer kappa of 0.68. Pervez et al., [11] found an interobserver kappa of 0.42 and intra-observer kappa of 0.33. Similar results were reported by Van Embden et al., [8] and Jin et al., [9] for the AO classification.

Other than when a clear separation exists between trochanteric area and intertrochanteric fractures (AO Type 3 or Evans Type Reverse), the degree of fracture comminution is difficult to determine. An analysis of the distribution showed that the population classified as A1 (simple) with conventional X-rays shifted into a large population classified as A2 (multifragment) with CT scanning for both observers. This study’s most important result serves to highlight the often complex nature of the fracture line in trochanteric fractures.

By using transverse axial CT slices and three-dimensional reconstruction, this study revealed a high frequency of lateral wall fractures (one of three fractures) and the challenge associated with identifying these fractures on conventional X-rays. An intramedullary implant must be used when the lateral wall is involved [26].

Conclusion

This goal of this unique study was not to suggest that CT scanning be used to evaluate trochanteric fractures, but
to increase our understanding of these fractures. Transverse axial CT slices provided a more precise analysis of the various trochanteric fracture lines in elderly subjects, and revealed that comminution and lateral wall involvement are more common than previously reported. Three-dimensional reconstructions did not provide a significant diagnostic benefit. The two main trochanteric fracture classification systems currently in use had little or no reproducibility, independent of the imaging modality used.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

References