Usefulness and reliability of two- and three-dimensional computed tomography in patients older than 65 years with distal humerus fractures

A. Jacquot\textsuperscript{a}, N. Poussange\textsuperscript{b}, J.-L. Charrissoux\textsuperscript{c}, P. Clavert\textsuperscript{d}, L. Obert\textsuperscript{e}, L. Pidhorzf, F. Sirveaux\textsuperscript{a}, P. Mansat\textsuperscript{f,g}, T. Fabre\textsuperscript{b,*}

\textsuperscript{a} Service d’orthopédie-traumatologie, centre chirurgical E.-Galle, 49, rue Hermite, 54000 Nancy, France
\textsuperscript{b} Service d’orthopédie-traumatologie, place Amélie-Raba-Leon, 33076 Bordeaux cedex, France
\textsuperscript{c} Département d’orthopédie-traumatologie, CHU Dupuytren, 2, avenue Martin-Luther-King, 87042 Limoges cedex, France
\textsuperscript{d} Centre de chirurgie orthopédique et de la main, 10, avenue Achille-Baumann, 67400 IllkirchGraffenstaden, France
\textsuperscript{e} Chirurgie orthopédique, traumatologique et plastique, centre hospitalier de Besançon, 2, boulevard Fleming, 25030 Besançon, France
\textsuperscript{f} Service de traumatologie, chirurgie orthopédique, centre hospitalier, 194, avenue Bubillard, 72037 Le Mans, France
\textsuperscript{g} Institut de l’appareil locomoteur, CHU de Toulouse, place du Dr-Baylac, 31059 Toulouse, France

Article Info

Article history:
Accepted 21 January 2014

Keywords:
Distal humeral fracture
Computed tomography
Inter-observer reliability

Abstract

Background: Distal humerus fractures are difficult to characterise and to classify according to the AO system. In this multicentre study, our objectives were to assess the usefulness of computed tomography (CT) and to measure intra-observer and inter-observer reliability according to observer experience.

Materials and methods: An online survey of professional practice was performed using a questionnaire based on a clinical case. Participants were asked to determine the AO classification using radiographs then to reappraise their answers after the addition of CT images. For the reliability study, 16 observers in five centres evaluated radiographs and CT scans of 26 distal humerus fractures. They used the radiographs to determine the AO classification and assess the main fracture characteristics then reappraised their findings after adding the CT images. The radiographs and 2D CT images were read twice at an interval of 2 weeks, and during the second reading, 3D CT images were available also. At least 1 month later, the same observers performed similar readings 2 weeks apart (radiographs and 2D CT images at the first reading and addition of 3D CT images at the second reading).

Results: Correct fracture classification was achieved in 95% of cases with the CT images compared to only 73% with the radiographs. CT led to diagnostic and therapeutic changes in 90% and 25% of cases, respectively. Inter-observer reliability was poor for both AO classification and fracture characteristics, not only with the radiographs and 2D CT images, but also with the added 3D CT images. In contrast, intra-observer reliability improved after the addition of 3D CT images. Assessment accuracy was influenced by image quality and geographic origin of the observer but not by observer experience.

Conclusion: CT improves diagnostic accuracy and, in some cases, changes the surgical strategy. In our study of a large number of observers, CT did not improve inter-observer agreement about the study variables. Intra-observer agreement was improved by 3D CT but not by 2D CT. Accuracy was not influenced by years of observer experience but was dependent on image quality, proficiency with computer-based tools and, above all, image observation and interpretation.

Level of evidence: Level III.

© 2014 Elsevier Masson SAS. All rights reserved.

Distal humerus fractures account for 5% of all fractures in patients older than 60 years of age [1]. The frequency of distal humerus fractures is predicted to increase 3-fold over the next 20 years [2–4].

The imaging studies performed in the emergency setting are often of limited quality. The result is poor reliability of fracture characterisation and classification, which hinders comparisons of published case-series studies [5]. Computed tomography (CT)-based imaging is available in most emergency centres and provides...
1. Material and method

For the practice survey, a questionnaire based on a clinical case was put on a website (Limesurvey: http://www.limesurvey.org). The patient was a 78-year-old self-sufficient woman who lived at home and had an unremarkable medical history. She had a fracture of the right distal humerus with no vessel or nerve injury. Antero-posterior and lateral radiographs of the elbow were available (Fig. 1).

The first part of the online questionnaire collected data on surgeon experience, AO fracture classification [10], and whether additional imaging studies were needed. In the second part of the questionnaire, a CT view in the coronal plane and two 3D CT reconstructions were shown (Fig. 2). Participants were asked to reappraise the AO classification based on these additional images.

For the multicentre reliability study, we invited five hospitals to participate. These five centres contributed a total of 26 sets of imaging studies inpatients older than 65 years of age with distal humerus fractures. Each set comprised antero-posterior and lateral radiographs of the elbow, thin-slice (<1.25 mm) 2D CT images, and multiplanar 3D reconstructions. At the emergency department, a whole-body CT scan was performed in 8 patients and a CT scan centred on the elbow in the remaining 18 patients. 3D reconstructions were obtained directly, either from the native images in 20 patients or from images subjected to pre-processing, notably using bone filters. The images showed all three elbow-joint bones, without subtraction of the proximal portions of the two forearm bones. Osirix 32b® software was used to analyse the images in DICOM format [11]. The 16 independent observers who participated in the study fell into three groups: 5 were senior residents, 5 were clinical fellows, and 6 were senior surgeons with more than 10 years of surgical traumatology practice. The observers were asked to classify each fracture in one of the nine AO system groups (A 1 to 3, B 1 to 3; and C 1 to 3). Diagrams of the AO classification with descriptions of each fracture group and subgroup taken from the original publications were available throughout the evaluation [10]. The participants were asked to assess the following fracture characteristics: articular comminution, metaphyseal comminution, fracture line in the coronal plane, strictly intra-articular fracture, and articular surface comminution. They were also asked to recommend a surgical strategy among the following: non-surgical treatment, internal fixation with its type (isolated screw fixation, one plate, or two plates), joint replacement surgery, and addition of a bone graft. Finally, the participants rated image quality as inadequate, acceptable, or optimal.

All 16 observers evaluated all 26 fractures, using the radiographs and 2D CT images. At least 2 weeks later, they re-evaluated the same fractures using not only the radiographs and 2D CT images, but also the 3D reconstructions. At least 1 month later, the observers re-evaluated the 26 fractures, using the radiographs and 2D CT images first then 2 weeks later, the radiographs and the 2D and 3D CT images.

Statistically, the kappa coefficient measures agreement among observers, after correction for the effect of chance [12,13]. The kappa coefficient provides information on inter-observer and intra-observer reliability for each study variable. We measured inter-observer reliability based on the first set of two evaluations and intra-observer reliability based in the second set of two evaluations. The kappa coefficient values were categorised as follows [13]: 0.00 to 0.20, slight agreement; 0.21 to 0.40, fair agreement; 0.41 to 0.60, moderate agreement; 0.61 to 0.80, substantial agreement and 0.81 to 1.00, almost perfect agreement. A Kappa coefficient value lower than 0 indicates complete disagreement and a value of 1 complete agreement. We computed the percentage of concordant
replies between the two readings to determine the impact of 3D CT images on the therapeutic management, that is, the differences in treatment recommendations between the first reading, with the radiographs and 2D CT images, and the second reading, with addition of the 3D CT images. The data were entered and analysed using Excel™. Statistical tests were performed using Statview™ software, with the chi-square test for qualitative variables, at the 5% significance level.

2. Results

We obtained 614 assessable questionnaires, completed by 64 residents, 60 clinical fellows, 123 orthopaedic surgeons with less than 10 years of experience, and 367 orthopaedic surgeons with more than 10 years of experience.

The fracture type was C1 in the AO classification. Based on the standard radiographs only, the fracture was classified correctly
by 73% of respondents overall, 65.6% of residents, 73.3% of clinical fellows, 73.2% of orthopaedic surgeons with less than 10 years of experience, and 73.5% of orthopaedic surgeons with more than 10 years of experience. The fracture was mistakenly classified as A2, C2, or another type by 14%, 8%, and 5% of respondents respectively. Surgeon experience did not significantly influence diagnostic accuracy ($P > 0.05$). After reading the CT images, 95% of respondents correctly classified the fracture as C1, a 22% increase compared to the standard radiographs alone. Among respondents who misclassified the fracture based on the radiographs alone, 90% determined the correct fracture type and 25% changed their therapeutic strategy after reading the CT images. Only 52% of respondents ordered CT as a complement to the radiographs. CT was ordered by 78% of the residents, 67% of the clinical fellows, 54% of the orthopaedic surgeons with less than 10 years of experience, and 44% of the orthopaedic surgeons with more than 10 years of experience (Table 1).

The evaluation of intra-observer reliability showed strong agreement, with kappa coefficients greater than 0.6 for over 60% of observers in the assessments of the AO classification and of the fracture characteristics using the 3D CT images. Intra-observer agreement was moderate (kappa < 0.6) when only the radiographs and 2D CT images were used (Fig. 3). Inter-observer reliability was poor for the AO classification and did not improve after addition of the 3D CT images (Fig. 4). For both the AO classification and assessment of fracture characteristics, observer experience had no significant effect on inter-observer reliability, which remained poor with the 2D CT images and after adding the 3D CT images (Figs. 5 and 6). A centre effect was noted, with agreement regarding the AO classification being good or almost perfect in two centres, moderate in one centre, and fair in two centres. Adding the 3D CT images had no substantial effect (Fig. 7). Better image quality was associated with higher kappa coefficients, which increased from the fair to the moderate range when image quality was rated as optimal (Fig. 8). Image quality was rated as inadequate for all 2D and 3D reconstruction images derived from whole-body CT scan data.

### Table 1

Responses to the questionnaire according to the level of experience of the respondents.

<table>
<thead>
<tr>
<th></th>
<th>Residents</th>
<th>Clinical fellows</th>
<th>Orthopaedic surgeon &lt; 10 years of activity</th>
<th>Orthopaedic surgeon &gt; 10 years of activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct radiological diagnosis (%)</td>
<td>65.6</td>
<td>73.3</td>
<td>73.2</td>
<td>73.5</td>
</tr>
<tr>
<td>CT ordered (%)</td>
<td>78</td>
<td>67</td>
<td>54</td>
<td>44</td>
</tr>
</tbody>
</table>

CT: computed tomography.

---

**Fig. 3.** Percentages of observers with intra-observer kappa coefficients greater than 0.6 (substantial or almost perfect agreement) in the evaluations of AO fracture classification and fracture characteristics using the 2D and 3D computed tomography images.

**Fig. 4.** Inter-observer kappa coefficient values for classification in the nine-class AO system using 2D and 3D images.

**Fig. 5.** Inter-observer kappa coefficient values for AO classification using 2D and 3D images according to observer experience.

**Fig. 6.** Inter-observer kappa coefficient values for analysis of the fracture characteristics using 2D and 3D images according to observer experience.
In our practice survey, less than three-fourths of respondents determined the correct AO classification based on the standard radiographs, and the level of experience had no influence on this parameter. The intercondylar fracture line was missed by 14% of respondents, who therefore mistakenly classified the fracture as A2 instead of C1. Importantly, we chose an optimal clinical case: the fracture was simple, without comminution, and radiograph image quality was very acceptable, which is not always the case in everyday practice. Radiographic diagnostic precision may be noticeably lower with complex fractures and poorer image quality. Reading the CT images resulted in correct fracture classification by 90% of the respondents who initially indicated the wrong fracture type. We believe CT is now an indispensable component of the initial pre-operative imaging armamentarium for most articular fractures, particularly in patients older than 60 years of age, fractures in which the main line is in the coronal plane, and comminuted fractures [14–16]. The available level of evidence supporting the use of CT is low [17], and in our study, over half the respondents did not order pre-operative CT imaging. The use of CT was dependent on surgeon age and, therefore, on surgeon experience. Greater surgeon experience was not associated with improved accuracy of the standard radiograph evaluation. Other parameters, such as usual practice patterns and availability of CT, influence the decision to obtain CT. Some surgeons continue to rely on fluoroscopy during upper limb traction under general anaesthesia in the operating room to evaluate the fracture and determine the treatment strategy [18]. This approach should be reserved for highly experienced trauma surgeons and should not be recommended or taught to young surgeons, among whom 78% ordered CT in our study. Among the respondents (22%) who corrected their diagnosis after reading the CT images, one-fourth changed their treatment strategy. This finding highlights the usefulness of CT for optimal prediction of the best surgical approach and of the equipment that will be needed in the operating room [18].

We report the first evidence of reliability regarding the evaluation of distal humerus fractures by surgeons with different degrees of experience working in different centres. We found a marked centre effect. Similarly, a study of inter-observer reliability for the Mason classification modified by Broberg and Morrey, as assessed using 2D CT images, showed differences in kappa coefficient values between surgeons in the US and in Europe [19]. These findings illustrate the challenges raised by comparing published case-series studies of distal humerus fractures conducted in different centres. In contrast, in our study, a greater number of years of observer experience did not improve inter-observer reliability, even in centres where the kappa coefficient was greater than 0.60. The conclusions drawn from CT images depend more heavily on personal interpretation of the data than on duration of practical experience but nevertheless strongly influenced by the level of training and experience acquired by the team in analysing the study criteria [8]. Two multicentre studies of distal radius and coronal fractures involved surgeons having varying levels of experience and can therefore be compared to our study [19,20]. In both studies, the number of observers was high and inter-observer reliability was fair and moderate for 2D and 3D CT images, respectively, regarding fracture classification and/or characteristics. In contrast, in studies of small numbers of observers (less than 5), particularly when the observers and images came from a single centre, inter-observer reliability was considered ‘good’ for the various study criteria in investigations of acetabular, distal radius, and proximal humerus fractures [6,21,22]. These data indicate that, in addition to the personal interpretation of each observer, factors such as strong familiarity with specific classifications, frequency of use, and proficiency with computer-based tools also play a role.

The use of 3D CT images improved intra-observer reliability for fracture classification and the assessment of fracture characteristics. These data are consistent with the results of single-centre studies of distal humerus fractures involving small numbers of observers [8,9]. The additional information supplied by 3D imaging seems to improve the analysis of the fracture by a given observer without improving agreement among observers. This discrepancy between inter-observer and intra-observer data highlights the difficulties raised by comparing our results based on an international classification that is well validated and viewed as the reference standard, since the analysis of the specific fracture characteristics did not improve inter-observer or intra-observer reliability.

The quality of 3D CT images is considerably better when the reconstructions are derived from native images. This method deserves to be used routinely, regardless of the joint under study, and will undoubtedly improve the inter-observer kappa coefficient values for fracture classification or the analysis of fracture characteristics.

A limitation of our study is the use of images obtained at different centres under a variety of imaging conditions, with some images being derived from whole-body CT scans or pre-processed electronically using filters. Furthermore, observer experience differed regarding use of the image processing software. CT of the elbow performed on an emergent basis is often obtained in a non-standardised position in terms of both flexion-extension and pronation-supination. It is of the utmost importance to use image processing software such as Osirix® to obtain multiplanar reconstructions located in a rigorously defined plane (relative to the planes of the joint and long bone axes). Finally, for our study, none of the 3D CT images were obtained after subtraction of the proximal portions of the two forearm bones, an image processing method that does not seem to improve inter-observer reliability [8]. All these limitations reflect the conditions of real-life practice in emergency trauma departments.
Our results confirm the usefulness of CT for evaluating and managing distal humerus fractures. CT imaging improves diagnostic precision and can therefore modify surgical decisions. To optimise image quality, the scan should be centred on the elbow and the 3D reconstruction images should be routinely obtained from the native images. Use of image reconstruction software capable of providing multiplanar images is indispensable. Regarding fracture classification and the analysis of fracture characteristics, 2D and 3D imaging did not improve inter-observer agreement in our study. In contrast, 3D imaging increased intra-observer agreement. Rather than the number of years of experience of the observers, image quality, use in everyday practice, proficiency in using computer-based tools and, above all, observer-specific interpretation of the images influenced the results of our reliability assessment.

Disclosure of interest

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

Acknowledgements

We thank the residents, clinical fellows, and orthopaedic surgeons of the hospitals in Besançon, Bordeaux, Nancy, Strasbourg, and Toulouse (France), who participated in this study.

References