Interobserver reliability of the AO system of classification of fractures of the distal radius was assessed using plain radiographs and CT. Five observers classified 30 Colles'-type fractures using only plain radiographs; two months later they were reclassified using CT in addition.

Interobserver reliability was poor in both series when detailed classification was used. By reducing the categories to five, interobserver reliability was slightly improved, but was still poor. When only two AO types were used, the reliability was moderate using plain radiographs and good to excellent with the addition of CT.

The use of CT as well as plain radiographs brings interobserver reliability to a good level in assessment of the presence or absence of articular involvement, but is otherwise of minor value in improving the interobserver reliability of the AO system of classification of fractures of the distal radius.

Received 31 October 1997; Accepted after revision 22 January 1998

The AO/ASIF classification system of fractures is often used in clinical studies of fractures of the distal radius to describe the severity of the injury.\(^1,2\) The interobserver reliability and reproducibility and clinical usefulness of the AO classification system have been criticised.\(^3-5\) Detailed AO classification of these fractures requires a good knowledge of the fracture lines entering the radiocarpal and radio-ulnar joints. Plain radiographs may not be sufficiently accurate to determine the exact morphology of the fracture and CT with image reconstruction is superior in this respect.\(^6,7\)

We have studied whether the use of CT in addition to plain radiographs improves the interobserver reliability of the AO classification system of fractures of the distal radius.

Patients and Methods

The 30 patients who were treated for Colles'-type fractures of the distal radius in Oulu University Hospital during 1996 were examined by using CT in addition to plain radiographs after closed reduction and immobilisation of the fracture in a cast. Anteroposterior and lateral plain radiographs taken before and after reduction, and CT axial scans and two-dimensional reformats in the sagittal and coronal planes were used for detailed AO classification,\(^8\) modified as described by Fernandez and Geissler\(^9\) (Fig. 1).

This system gives 14 possible categories for a Colles'-type fracture. For statistical evaluation, the 14 categories were first reduced to five main classes (A2, A3, C1, C2, C3) and finally to two main types (A and C). This was possible because classification in the AO system is done sequentially from main types to main classes and subclasses. The radiographs were reviewed by five assessors: an orthopaedic registrar, two hand surgeons and two radiologists who were familiar with the classification system, but who had not used it in clinical practice. Each assessor read the instructions on the classification described by Müller et al.,\(^8\) a written description and a scheme of the AO classification, and attended a short lecture on the principles of the classification.

The assessors first classified the fractures using only plain radiographs. Two months later the fractures were reclassified using CT in addition to the plain radiographs, with no feedback from the previous classification. The radiological material available for classification was exactly the same for all assessors on both occasions. All radiographs were reviewed independently.

Statistical methods. Kappa statistics and the percentage agreement (the percentage of assessors who agreed with one another) were the statistical methods used. The kappa coefficient for multiple raters was calculated according to the method described by Fleiss.\(^10\) Percentage agreement,
although involving the possibility that some agreement will occur purely by chance, gives an overview of agreement between two assessors. Interpretation of the kappa values was carried out according to the guidelines proposed by Fleiss, and by Landis and Koch. According to Fleiss, kappa values less than 0.40 indicate poor reliability, 0.40 to 0.75 fair to good reliability and 0.75 to 1.0 excellent reliability. According to Landis and Koch, kappa values of 0.00 to 0.20 represent slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement and 0.61 to 0.80 substantial agreement. A value above 0.80 is considered almost perfect agreement.

Results

The percentage agreement when using the complete AO classification between different pairs of assessors ranged from 17% to 40% using plain radiographs and from 17% to 50% with the addition of CT. The overall kappa coefficients were 0.18 for plain radiographs and 0.16 for CT, indicating poor interobserver reliability in both cases. After reducing the categories to five main classes, the percentage agreement between assessors ranged from 23% to 50% using plain radiographs and from 20% to 63% with the addition of CT. The respective overall kappa coefficients were 0.23 and 0.25 showing a slight increase in reliability after adding CT, but the result was still regarded as only poor to fair. After reducing the classes to two main types, the percentage agreement varied from 60% to 87% for plain radiographs and from 87% to 100% for CT. The overall kappa coefficients were 0.48 (moderate) for plain radiographs and 0.78 (good to excellent) with CT.

Discussion

Interobserver reliability is essential for any system of classification. We found this to be poor when using the complete AO classification from plain radiographs. The addition of CT assessment did not improve reliability. Reducing the categories to five main classes still produced no improvement. Only after restricting classification to two types (extra-articular and intra-articular) did agreement reach an acceptable level by using CT in addition to plain radiographs. By then, however, most of the information from the classification system had been lost. CT with two-dimensional reformats helped in assessing the presence or absence of intra-articular fracture lines, but difficulties arose increasingly when the exact morphology of the fracture was to be

Fig. 1

Modified AO classification of fractures of the distal radius.

VOL. 80-B, NO. 4, JULY 1998
classified. The degree of metaphyseal comminution was more difficult to judge from CT scans than from plain radiographs, and assessing the fractures as simple or comminuted appeared to be arbitrary and difficult.

Our results agree with those published by Andersen et al, and Kreder et al, who found acceptable reproducibility of the AO classification of fractures of the distal radius using plain radiographs only for the main types. The presence or absence of articular displacement can be classified with high consistency from plain radiographs. Colles'-type fractures are covered by part of the AO classification system of fractures of the distal radius and simple articular injuries (type B) were not involved in this series. Our results are not therefore directly comparable with those published previously. Poor interobserver reliability and intraobserver reproducibility have also been reported in the classifications of Frykman, Melone and Mayo using only plain radiographs. Older’s classification, which is based mainly on the shortening of the radius and not on the morphology of the fracture, has been shown to have good reproducibility and correlates well with the radiological and clinical outcome.

To our knowledge CT has not been used before in the evaluation of the interobserver reliability of classification of fractures of the distal radius. Johnston et al in a series of 19 patients examined by CT in addition to plain radiographs, reported that in five CT indicated higher Frykman classes than did plain radiographs. Classification was carried out by different assessors. They did not assess reproducibility of the classification. We also found that assessors tended to decide on higher AO classes when CT scans were available for classification.

Unsatisfactory interobserver reliability does not arise only when considering fractures of the distal radius. Similar problems have been described with the Lauge-Hansen classification in ankle fractures, the Neer and AO classifications in fracture of the proximal humerus and with the Garden classification of fractures of the femoral neck. CT does not improve either intraobserver reproducibility or interobserver reliability of the Neer and AO classifications in fractures of the proximal humerus.

An ideal clinical classification of fractures would consider the severity of the injury, serve as a basis for recommendations for treatment and assist in the evaluation of the results. It may be impossible to classify the severity of a fracture based only on the bone lesion, since with fractures of the distal radius the patients’ age and medical condition may often be more important than the injury itself when choosing the method of treatment. A purely descriptive classification of fractures may be useful in everyday communication, but interobserver reliability is essential.

We have attempted to assess the clinical application of the AO classification system and not to assess its scientific validity. The complete classification when applied to a Colles’ fracture shows poor interobserver reliability and only the main groups can be used reliably to grade the severity of the lesion.

The authors would like to thank Mr Pentti Nieminen, PhD, for his assistance in statistical evaluation of the results. No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article.

References