Complex distal humerus trauma

Brouwer, K.M.

Citation for published version (APA):
CHAPTER 5
Diagnostic Accuracy of Two-Dimension and Three-Dimensional Imaging and Modeling of Distal Humerus Fractures
Brouwer KM, Lindenhovius AL, Dyer GS, Zurakowski D, Mudgal CS, Ring D
*In Press, J Shoulder Elbow Surg*
Abstract

Purpose: This investigation used prospectively recorded intra-operative evaluation as the reference standard for distal humerus fracture type and characteristics in order to measure the diagnostic performance characteristics of computed tomography and physical models. In secondary analyses we assessed the reliability of classification.

Methods: Thirty-five fractures were evaluated by the treating surgeon and first assistant on radiographs and two-dimensional (2D) computed tomography (CT) images first; a second time based on radiographs, 2D and three dimensional (3D) CT images; a third time based on 2D and 3D-CT, as well as 3D physical models; and a fourth time based on intra-operative visualization of the fracture characteristics. The intra-operative evaluation of the attending surgeon was used as the reference standard.

Results: The addition of 3DCT and the 3D models to 2DCT and radiographs led to significant improvements in sensitivity, but not specificity, in the diagnosis and proposed treatment, and improved the interobserver agreement with respect to specific fracture characteristics but not classification.

Conclusions: Increasingly sophisticated imaging and modeling leads to slight but significant improvements in diagnostic performance characteristics and interobserver agreement on fracture characteristics.

Type of study/level of evidence: Diagnostic, Level 2
Introduction

Three-dimensional imaging seems more intuitive because structures look similar to what the surgeon sees in the operating room. Computer generated bone models have been successfully used in the planning of surgery to address malunion of the distal radius or humerus fractures\textsuperscript{1-7} and might be even more intuitive for understanding the injury and planning treatment.

In one study, three-dimensional computed tomography (3DCT) of fractures of the distal humerus improved both the intraobserver and the interobserver reliability of the AO classification system, but only improved the intraobserver agreement for fracture characteristics and treatment recommendations when compared to two-dimensional computed tomography (2DCT). In general, more sophisticated imaging seems to affect intraobserver agreement more than interobserver agreement.\textsuperscript{8} Furthermore, 3DCT did not improve the diagnostic characteristics (sensitivity, specificity, and accuracy) for the recognition of specific fracture characteristics, but that study used a retrospective, medical record based reference standard.\textsuperscript{8}

This investigation used prospectively recorded intra-operative evaluation as the reference standard for fracture type and characteristics. Diagnostic performance characteristics were calculated for 2D and 3DCT and 3D models with respect to this reference standard to test our hypothesis that 3D-CT images and models improve the accuracy of classification and diagnosis of specific distal humerus fracture characteristics. In secondary analyses, the agreement between surgeon and first assistant on the classification and characterization of these fractures was assessed.

Material and Methods

\textit{Inclusion and Exclusion Criteria}

Between 2007 and 2010 patients with a distal humerus fracture treated at two Level 1 trauma centers were invited to enroll in a prospective cohort study approved by our Human Research Committee. Inclusion criteria were 1) fracture of the distal humerus; 2) election of operative treatment; 3) Computed Tomography scan (CT) of sufficient quality to make 3D images and models; 4) age 18 years or older. Exclusion criteria were pregnant women and patients unable to give informed consent. Thirty-five patients satisfied the inclusion and exclusion criteria.

There were 17 men (49%) and 18 women with a mean age of 50 years (range, 20 to 94 years). The distal humerus fracture was an isolated injury in 21 patients (60%) and was associated with open wounds, dislocation or fracture of the radial head, coronoid process or olecranon in 14 patients. The left side was injured in 27 patients (77%) and the right side in 8 patients (23%). Twenty patients (57%) fractured their elbow in a fall from a standing height, 4 (11%) from a greater height, 4 (11%) in a motor vehicle collision (MVC), 5 (14%) in a sports related injury and 2 (6%) were gunshot fractures.
Three patients (9%) had both an open wound and an ulnohumeral dislocation. One of these patients also had a coronoid and a radial head fracture. Three patients (9%) had an open wound without dislocation. Two of these patients had associated fractures: one patient had a radial head, a coronoid and an olecranon fracture; the other patient had a non-displaced olecranon fracture with intra-articular extension.

Four patients (11%) had an ulnohumeral dislocation without an open wound. Three of these had associated fractures: one patient had a coronoid fracture, one patient had a coronoid and radial head fracture; the third patient had a coronoid fracture and an olecranon fracture. Four patients (11%) had associated fractures, without dislocation or open wound. There were 2 patients with a radial head fracture. One patient had coronoid fracture and one patient had both a radial head and a fracture of the ulnar diaphysis.

**Evaluation**

Three-dimensional CT reconstructions with the radius and ulna subtracted were made. The DICOM (Digital Imaging and Communications in Medicine) files from the Computed Tomography scans were sent to Medical Modeling LLC (Golden, CO) for manufacture of the 3D physical model reconstructions.

The treating surgeons and their first assistant classified the fracture according to the Comprehensive Classification of Fractures, diagnosed the presence or absence of several fracture characteristics, and indicated treatment plans. These ratings were completed four times: initially based upon radiographs and 2D images alone; a second time based on radiographs, 2D and 3D-CT images; a third time on radiographs, 2D, 3D-CT and 3D physical models; and a fourth time based on intra-operative visualization of the fracture characteristics. The fourth questionnaire completed by the intra-operative evaluation of the surgeon represented the reference standard.

**Fracture Characteristics and Treatment plan:**

Each observer was asked diagnose the presence or absence of the following fracture characteristics: 1) coronal fracture line; 2) more than 3 articular fragments; 3) metaphyseal comminution; 4) entirely separated articular fragments; 5) impaction (stable fracture) of the articular surface. Proposed treatment was evaluated by choosing one or more of the following five treatment options: standard plating, plating variation like a third plate or parallel plate, small headless screws or buried fixation, bone graft; and need to be prepared for total elbow arthroplasty.

**Statistical Analysis**

We used a chance-corrected kappa ($\kappa$) coefficient with strength of agreement assessed with the benchmarks of Landis and Koch\(^9\) to measure the interobserver agreement concerning fracture class, fracture characteristics and treatment proposal was measured for each method (2D, 2D/3D, 2D/3D with physical model, direct intraoperative view).
In order to account for the same 35 cases evaluated by multiple surgeons using each of 4 different methods (2D, 2D/3D, 2D/3D with physical model, direct intraoperative view) logistic regression was then applied using a generalized estimating equations (GEE) strategy with a binomial distribution for binary yes/no fracture characteristics and a multinomial logit distribution for fracture classification.

The AO classification was considered to the Group level (3 groups for each of 3 types for a total of 9 options). Differences between the methods were determined using the maximum likelihood Wald chi-square test with a two-tailed p < 0.05 as the criterion for statistical significance. According to power analysis performed prior to initiating the study, 30 fractures provide 80% power (\(\alpha = 0.05, \beta = 0.20\)) to detect significant inter-observer agreement using the kappa coefficient as well as in comparing diagnostic characteristics between the two imaging modalities. Sensitivity, specificity, and accuracy were calculated using standard formulas with the intra operative findings of the attending surgeon as the reference standard. McNemar’s test for paired binary data was used to test for statistically significant differences.

Results
The addition of 3DCT and the 3D models to 2DCT and radiographs led to significant improvements in sensitivity, but not specificity, in the diagnosis of coronal fracture line, more than 3 articulating fragments, metaphyseal comminution, separated articular fragments, impaction and proposed treatment (Table 1). More sophisticated imaging improved the interobserver agreement of diagnosis of, coronal fracture line, more than 3 articulating fragments, metaphyseal comminution, separated articular fragment, and proposed treatment; however, classification did not improve with more sophisticated imaging. (Table 2)

Discussion
The strengths of this investigation include the prospective design, adequate power, and the intra-operative reference standard. With prospective direct intraoperative view as the gold standard, 3D-CT and 3D models achieve slightly, but significantly higher diagnostic performance characteristics than radiographs or two-dimensional computed tomography alone.

This study should be interpreted in light of the fact that we were unable to consistently get observations completed before surgery (in part due to the inherent delay in receiving the physical 3D model), so that ratings of the radiological images were—in essence—sometimes retrospective and biased by having seen the operative exposure. There is probably some spectrum bias although the majority of distal humerus fractures at our institution are imaged with computed tomography. When reviewing these data, one should also bear in mind that the second observer was either a resident or fellow, so that observer variability may largely reflect differences in training and experience. Prior studies on acetabular, proximal humerus, distal radius, and coronoid fractures have demonstrated improved agreement in characterization and
classification of fractures with 3DCT compared to 2DCT and radiographs alone, as well as increased intra-observer agreement, more so than inter-observer agreement.\textsuperscript{12-21}

Other retrospective work on the distal radius\textsuperscript{22} suggested that three-dimensional computed tomography can improve the intra-observer reliability of diagnosis of fracture characteristics, but found a more limited influence interobserver agreement when compared to 2D imaging alone. In other studies it was found that 3D CT reconstructions improve interobserver agreement on classification and treatment of coronoid fractures when compared with 2D CT\textsuperscript{18}; that 3D CT reconstructions significantly improved the interobserver agreement on classification of proximal humerus fractures\textsuperscript{23}; and that three-dimensional computed tomography is the most accurate method to assess true displacement in medial humeral epicondyle fractures in children.\textsuperscript{24} Some studies, such as one looking at tibial plateau fractures found that 3D CT did not improve the reliability of classification or characterization compared to 2D CT alone.\textsuperscript{25}

There is still a question as to whether reliability improves with more sophisticated imaging simply because the observer has more data as an increasing number of studies are reviewed.\textsuperscript{26} It has also been suggested that interobserver variability may be a reflection of the level of training, the awareness and understanding of fracture concepts and variations in opinion rather than the nature or quality of images.

**Conclusion**

More sophisticated imaging, three-dimensional computed tomography and three-dimensional printed models, in distal humerus fractures achieve higher diagnostic performance characteristics than radiographs or two-dimensional computed tomography alone.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coronal fracture line</td>
<td>80 (65-89)</td>
<td>87 (70-94)</td>
<td>85 (71-93)</td>
<td>77 (59-88)</td>
<td>93 * (80-97)</td>
<td>80 (63-90)</td>
</tr>
<tr>
<td>&gt;3 articular fragments</td>
<td>75 (59-86)</td>
<td>82 (66-92)</td>
<td>86 * (71-93)</td>
<td>88 (73-95)</td>
<td>78 (61-88)</td>
<td>88 (73-95)</td>
</tr>
<tr>
<td>Metaphyseal comminution</td>
<td>62 (45-76)</td>
<td>86 (71-94)</td>
<td>76 (60-88)</td>
<td>92 (78-97)</td>
<td>83 * (69-91)</td>
<td>94 (82-98)</td>
</tr>
<tr>
<td>Separated articular fragments</td>
<td>78 (65-86)</td>
<td>75 (47-91)</td>
<td>86 * (75-93)</td>
<td>83 (55-95)</td>
<td>86 * (75-93)</td>
<td>83 (55-95)</td>
</tr>
<tr>
<td>Impaction</td>
<td>43 (28-58)</td>
<td>83 (67-93)</td>
<td>65 (50-78)</td>
<td>77 (60-88)</td>
<td>73 * (57-84)</td>
<td>77 (59-88)</td>
</tr>
</tbody>
</table>

**Proposed Treatment**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard plating techniques</td>
<td>89 (67-97)</td>
<td>71 (58-82)</td>
<td>100 (82-100)</td>
<td>79 (66-88)</td>
<td>100 (82-100)</td>
<td>77 (64-86)</td>
</tr>
<tr>
<td>Third plate/parallel plates</td>
<td>60 (42-75)</td>
<td>98 (87-100)</td>
<td>83 (66-93)</td>
<td>95 (84-99)</td>
<td>77 * (59-88)</td>
<td>98 (87-100)</td>
</tr>
<tr>
<td>Headless screws/buried fixation</td>
<td>81 (68-90)</td>
<td>68 (47-84)</td>
<td>88 * (75-94)</td>
<td>64 (43-80)</td>
<td>88 * (75-94)</td>
<td>82 (61-93)</td>
</tr>
<tr>
<td>May need bone graft</td>
<td>72 (49-88)</td>
<td>79 (66-88)</td>
<td>83 * (61-94)</td>
<td>85 (72-92)</td>
<td>72 (50-88)</td>
<td>92 (81-97)</td>
</tr>
<tr>
<td>Prepared for total elbow</td>
<td>50 (15-85)</td>
<td>97 (90-99)</td>
<td>100 * (51-100)</td>
<td>92 (83-97)</td>
<td>100 * (51-100)</td>
<td>95 (87-98)</td>
</tr>
</tbody>
</table>

Values are percentages with 95% confidence interval based on attending surgeon for each method compared to intraoperative direct view gold standard (N = 35 paired cases).

* Statistically significant compared to sensitivity for 2D CT alone (all p < 0.05). There were no significant differences in specificity between the three methods.
<table>
<thead>
<tr>
<th>Fracture classification</th>
<th>Xray + 2D CT</th>
<th>With 3D CT</th>
<th>With 3D model</th>
<th>Intra-operative view (Gold Standard)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AO Classification</td>
<td>0.74†</td>
<td>0.78†</td>
<td>0.74†</td>
<td>0.64†</td>
</tr>
<tr>
<td>Fracture characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronal fracture line</td>
<td>0.43†</td>
<td>0.24</td>
<td>0.46†</td>
<td>0.53†</td>
</tr>
<tr>
<td>&gt; 3 articular fragments</td>
<td>0.49†</td>
<td>0.60†</td>
<td>0.67†</td>
<td>0.66†</td>
</tr>
<tr>
<td>Metaphyseal comminution</td>
<td>0.63†</td>
<td>0.71†</td>
<td>0.71†</td>
<td>0.66†</td>
</tr>
<tr>
<td>Separated articular fragments</td>
<td>0.47†</td>
<td>0.41*</td>
<td>0.41*</td>
<td>0.32*</td>
</tr>
<tr>
<td>Impaction</td>
<td>0.14</td>
<td>0.26</td>
<td>0.21</td>
<td>0.29</td>
</tr>
<tr>
<td>Proposed treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard plating techiques</td>
<td>0.40†</td>
<td>0.41†</td>
<td>0.38†</td>
<td>0.38†</td>
</tr>
<tr>
<td>Third plate/parallel plates</td>
<td>0.35*</td>
<td>0.58†</td>
<td>0.51†</td>
<td>0.50†</td>
</tr>
<tr>
<td>Headless screw s/buried fixation</td>
<td>0.62†</td>
<td>0.44†</td>
<td>0.62†</td>
<td>0.64†</td>
</tr>
<tr>
<td>May need bone graft</td>
<td>0.62†</td>
<td>0.42*</td>
<td>0.46†</td>
<td>0.58†</td>
</tr>
<tr>
<td>Prepared for total elbow</td>
<td>0.48†</td>
<td>0.62†</td>
<td>0.53†</td>
<td>1.00†</td>
</tr>
</tbody>
</table>

Data are kappa (k) values based on 35 cases evaluated by two independent surgeons. Guidelines for strength of observer agreement: k = 0.0-0.20 (slight), k = 0.21-0.40 (fair), k = 0.41-0.60 (moderate), k = 0.61-0.80 substantial, k = 0.81-1.00 almost perfect. Significant interobserver agreement (* p < 0.05; † p < 0.01).
References