Intraobserver and Interobserver Agreement in the Classification and Treatment of Midshaft Clavicle Fractures

Grant L. Jones,†‡ MD, Julie Y. Bishop,†‡ MD, Brian Lewis,§ MD, Angela D. Pedroza,† MPH, and MOON Shoulder Group||
Investigation performed at Wexner Medical Center, The Ohio State University, Columbus, Ohio, USA

Background: With the recent emphasis on performing open reduction and internal fixation on midshaft clavicle fractures with complete displacement, comminution, and >2 cm of shortening, it is important to determine the reliability of orthopaedic surgeons to assess these variables on standard plain radiographs and to determine the agreement among orthopaedic surgeons in choosing the treatment.

Purpose: To determine the intra- and interobserver reliability in the classification of midshaft clavicle fractures via standard plain radiographs and to determine the intra- and interobserver agreement in the treatment of these fractures.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: Charts of patients seen by the 2 senior authors from 2006 to 2011 were reviewed to identify patients treated for clavicle fractures (CPT codes 23500 and 23515). Anteroposterior and 30° cephalad radiographs were selected, representing midshaft clavicle fractures treated both operatively and nonoperatively. Thirty pairs of radiographs were included in the investigation. The radiographs were standardized for size to allow accurate measurements within a non-PACS (picture archiving and communications system) program, and a PDF document was created with all representative radiographs. Clinical scenarios were created for each set of radiographs, and the evaluators were asked to (1) measure the degree of shortening in millimeters, (2) determine the percentage displacement, (3) determine whether the fracture was comminuted, and (4) state whether they would treat the fracture operatively or nonoperatively. The radiographs, along with instructions on how to use the measuring tool with Adobe Reader, were distributed to 22 shoulder/sports medicine fellowship–trained orthopaedic surgeons, then reordered and redistributed approximately 3 months later. Sixteen surgeons completed 1 round of surveys, and 13 surgeons completed both rounds.

Results: Interrater agreement was moderate for displacement of 0%-49% (κ = 0.71, P < .001) and >100% (κ = 0.73, P < .001), with minimal agreement for displacement of 50%-100% (κ = 0.39, P < .001). There was moderate interrater agreement for the presence/absence of comminution (κ = 0.75, P < .001). Interrater agreement was weak for shortening of 0-5.0 mm (κ = 0.58, P < .001) and >30.0 mm (κ = 0.51, P < .001), with minimal agreement for shortening of 5.1-10.0 mm (κ = 0.22, P < .001) and no agreement for the other 4 categories. Interrater analysis showed weak agreement on whether surgical treatment was recommended (κ = 0.40, P < .001). Intrarater agreement was strong for comminution (κ = 0.80, P < .0001), moderate for both displacement (κ = 0.76, P < .001) and operative treatment (κ = 0.64, P < .001), and minimal for shortening (κ = 0.38, P < .001). The following variables statistically predicted whether surgery was recommended (P < .001): (1) the odds of surgery were 2.26 if comminution was noted, holding displacement and the interaction between displacement and shortening constant, and (2) the odds of surgery were 3.37 if there is displacement of >100% compared with displacement of 0%-49%, holding comminution and shortening constant.

Conclusion: Standard plain unilateral radiographs of the clavicle are insufficient to reliably determine the degree of shortening of clavicle fractures and the need for surgery among shoulder/sports medicine fellowship–trained orthopaedic surgeons. Consideration should be made to not use shortening as the sole determinant for whether to proceed with surgical intervention or to use other radiographic modalities to determine the amount of shortening.

Keywords: clavicle fracture; classification; agreement; treatment; radiographs

Traditionally, midshaft clavicle fractures have been treated nonoperatively. Neer16 and Rowe23 both reported low nonunion rates in large numbers of clavicular fractures treated nonoperatively: 0.1% and 0.8%, respectively.
Additionally, they reported higher nonunion rates after primary open reduction: 3.6% and 3.7%, respectively.\textsuperscript{16,23} In 1981, however, Zenni et al\textsuperscript{30} found 100% clinical and radiographic healing rate in 25 patients with comminuted and angulated clavicle fractures treated with open reduction and internal fixation with a large Kirschner wire or Steinmann pin, and the authors suggested that severely comminuted and angulated clavicle fractures should be managed with open reduction and internal fixation.

More recent studies evaluating the radiographic and functional outcome of nonoperatively treated clavicle fractures have demonstrated a higher rate of nonunion or symptomatic malunion in nonoperatively treated fractures than previously reported.\textsuperscript{5,6,8,10,12,13,18-22,26} Varying degrees of clavicular shortening, clavicular displacement, and comminution (particularly a vertical butterfly fragment) have been associated with poorer functional outcomes.\textsuperscript{5,6,8,10,12,13,18-22,26} Most of these investigations agree that significant displacement and comminution can lead to worse clinical outcomes; however, there is disagreement on the effect of clavicle shortening and on the degree of clavicle shortening (reported range from $>$15 to $>$23 mm of shortening) required to produce poorer outcomes.\textsuperscript{5,6,8,13,17-20,26}

Several recent evidence level 1-3 investigations have demonstrated improved radiographic and clinical outcomes in patients with displaced and comminuted clavicle fractures treated with open reduction and internal fixation versus nonsurgical management.\textsuperscript{1,4,9,11,15,28} A recent meta-analysis of randomized clinical trials\textsuperscript{14} and a recent systematic review of the literature\textsuperscript{27} found that operative treatment of substantially displaced clavicle fractures resulted in a significantly lower rate of nonunion and symptomatic malunion and an earlier return to function than nonoperative treatment. However, there was little evidence to support that the long-term functional outcome of operatively treated patients was significantly superior.\textsuperscript{14}

With the recent trend toward operating on displaced, shortened, and comminuted clavicle fractures based on the above studies, it is important to study the reliability between orthopaedic surgeons in determining the degree of displacement, degree of shortening, and the presence of comminution and assessing the need for surgery based on standard plain radiographs of the clavicle. The typical trauma radiographic series of the clavicle in the emergency room or the clinic setting includes anteroposterior (AP) and cephalic tilt views. The purposes of our study were (1) to determine the intra- and interobserver reliability in assessing the degree of shortening, degree of displacement, and the presence of comminution on AP and 30° cephalad radiographs of the clavicle and (2) to evaluate the intra- and interobserver reliability in determining whether the fracture should be treated operatively or nonoperatively between shoulder/sports medicine fellowship–trained orthopaedic surgeons.

**MATERIALS AND METHODS**

The study was approved by the institutional review board of Wexner Medical Center at The Ohio State University. Clavicle fracture radiographs were identified by searching the records of 2 shoulder surgeons at (G.L.J., J.Y.B.) from 2006 to 2011 via the International Classification of Diseases, 9th Revision (ICD-9) for the diagnosis of clavicle fracture (810.0). The medical record numbers were identified. Based on these medical record numbers, the picture archiving and communications system (PACS) archives at The Ohio State University (RADWeb) were reviewed. Patients who did not have AP and 30° cephalic tilt injury films in PACS were excluded. All included radiographs were assessed to be of good quality, and all the included radiographs were downloaded from the same system to limit variability. Radiographs of patients younger than 18 years or older than 89 years were excluded.

The remaining radiographs were reviewed and selected so that injuries across the entire spectrum of injury severity were included. Using an alpha value of 0.05 and a power of 0.9, we arrived at 30 clinical vignettes. We utilized the amount of shortening as the primary outcome measure when performing the sample size calculation. The radiographs were then de-identified. The images were transferred into PowerPoint format (Microsoft Inc, Redmond, Washington, USA) and resized on the basis of differences in image magnification to standardize measurements. The PowerPoint file was then converted into a PDF (Adobe Systems Inc, San Jose, California, USA) to use the calibrated measuring tool to determine the amount of shortening and degree of displacement. Each image was measured within the PDF and compared with measurements within the PACS system to confirm the accuracy of the conversion process.

The participants were members of a multicenter group that studies outcomes from shoulder surgery. The PDF file containing the radiographs and detailed instructions on how to use the measuring tool within Adobe Acrobat Reader (Adobe Systems Inc) and surveys were sent to the orthopaedic surgeons. In the surveys, the surgeons were asked to (1) measure and record the amount of shortening in millimeters to 1 decimal point, (2) determine the percentage displacement (0%-49%, 50%-99%, or 100%), (3) determine whether the fracture was comminuted (yes/no), and (4) determine whether each fracture should be.
treated operatively (yes/no). Clear instructions were given on how to determine and document the percentage displacement, the presence of comminution, and the amount of shortening to standardize these measurements. No specific instructions or criteria were given to the evaluating surgeons on how to determine the need for surgery. We instructed the surgeons to use the criteria that they would utilize in their own clinical practices to determine the need for surgery.

The senior authors were excluded from participation, which left 22 members who were sent the initial survey. The members were given 2 months to complete the survey. Four months after the receipt of the last first-round survey, the radiographs were placed in a different order, and the survey and radiographs were resent to the shoulder group members. Sixteen surgeons completed 1 round of surveys, and 13 physicians completed both rounds.

After the survey, the surgeons were asked to rank the importance of the following 3 variables in their determination of surgical versus nonsurgical treatment: amount of shortening, degree of displacement, and presence of comminution. They were also asked to select what factor or combination of factors was utilized to decide on surgical versus nonsurgical treatment: (1) amount of shortening only, (2) degree of displacement only, (3) presence of comminution only, (4) amount of shortening and degree of displacement, (5) amount of shortening and presence of comminution, or (6) amount of shortening, degree of displacement, and presence of comminution.

**Statistical Methods**

Inter- and intrarater agreement was determined for the shortening, displacement, comminution, and decision to operate by evaluating the kappa values: 0-0.20 = none, 0.21-0.39 = minimal, 0.40-0.59 = weak, 0.60-0.79 = moderate, 0.80-0.90 = strong, and >0.90 = almost perfect. The amount of shortening measured on radiograph was broken into 7 categories: 0-5 mm, 5.1-10.0 mm, 10.1-15.0 mm, 15.1-20.0 mm, 20.1-25.0 mm, 25.1-30 mm, and >30 mm. Displacement was broken into 3 categories: 0%-50%, 50%-99%, and >100%. Comminution and responses to the question “Would you treat this fracture operatively” were limited to yes/no.

Multivariable logistic regression was performed to determine if shortening, displacement, or comminution could accurately predict surgical intervention.

**RESULTS**

**Interrater Agreement**

Interrater agreement is summarized in Table 1 and was as follows. Raters showed moderate agreement for displacement of 0%-49% ($\kappa = 0.71, P < .001$) and >100% ($\kappa = 0.73, P < .001$) and minimal agreement for displacement of 50%-100% ($\kappa = 0.39, P < .001$). Moderate agreement was seen for the presence/absence of comminution ($\kappa = 0.75, P < .001$). Raters showed weak agreement for shortening of 0-5.0 mm ($\kappa = 0.58, P < .001$) and >30.0 mm ($\kappa = 0.51, P < .001$), minimal agreement for shortening of 5.1-10.0 mm ($\kappa = 0.22, P < .001$), and no agreement for the other 4 categories. Finally, raters showed weak agreement on whether surgical treatment would be recommended ($\kappa = 0.40, P < .001$).

**Intrarater Agreement**

Intrarater agreement was strong for comminution ($\kappa = 0.80, P < .001$), moderate for both displacement ($\kappa = 0.76, P < .001$) and operative treatment ($\kappa = 0.64, P < .001$), and minimal for shortening ($\kappa = 0.38, P < .001$) (Table 1).

**Multivariable Logistic Regression**

The odds of recommending surgery were 3.37 if there was displacement of >100% compared with displacement of 0%-49%, holding comminution and shortening constant ($P = .008$). The odds of recommending surgery were 2.26 if comminution was present compared with if no comminution was noted, holding displacement, shortening, and the interaction between displacement and shortening constant ($P = .001$).

**Survey of Variables Used to Determine Surgical Versus Nonsurgical Treatment**

All the evaluators ranked the presence of comminution as the least important factor in deciding on treatment. Sixty percent of our evaluators stated that the amount of shortening was the most important criterion, and the remaining 40% ranked the degree of displacement as the most important. Forty-six percent of our evaluators stated that they utilized both the amount of shortening and the degree of

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Intra- and Interobserver Reliability Results*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intraobserver Reliability ($P &lt; .001$)</td>
</tr>
<tr>
<td>Operative (Y/N)</td>
<td>0.64</td>
</tr>
<tr>
<td>Comminution (Y/N)</td>
<td>0.80</td>
</tr>
<tr>
<td>Displacement, %</td>
<td>0.76</td>
</tr>
<tr>
<td>0-49</td>
<td>0.71</td>
</tr>
<tr>
<td>50-100</td>
<td>0.39</td>
</tr>
<tr>
<td>&gt;100</td>
<td>0.75</td>
</tr>
<tr>
<td>Shortening, mm</td>
<td>0.38</td>
</tr>
<tr>
<td>0-5</td>
<td>0.58</td>
</tr>
<tr>
<td>5.1-10</td>
<td>0.22</td>
</tr>
<tr>
<td>10.1-15</td>
<td>0.10</td>
</tr>
<tr>
<td>15.1-20</td>
<td>0.08</td>
</tr>
<tr>
<td>20.1-25</td>
<td>0.12</td>
</tr>
<tr>
<td>25.1-30</td>
<td>0.11</td>
</tr>
<tr>
<td>&gt;30.1</td>
<td>0.51</td>
</tr>
</tbody>
</table>

*aLevel of agreement ($\kappa$): 0-0.20 = none, 0.21-0.39 = minimal, 0.40-0.59 = weak, 0.60-0.79 = moderate, 0.80-0.90 = strong, >0.90 = almost perfect.
displacement but not comminution when deciding on treatment, while 46% of the observers stated that they used all 3 factors when selecting a treatment.

**DISCUSSION**

Traditionally, midshaft clavicle fractures have been treated nonoperatively. However, more recent studies have raised concerns that certain midshaft clavicle fractures treated nonoperatively have a higher rate of nonunion or symptomatic malunion. In addition, several recent comparison studies have demonstrated improved union rates and improved functional outcome with operative treatment of displaced clavicle fractures versus nonoperative treatment.

Several studies of nonoperatively treated clavicle fractures have found variables such as clavicular shortening, clavicular displacement, and comminution to be associated with higher rates of nonunion or symptomatic malunion. The importance of shortening and the degree of shortening in determining outcomes is unclear in the literature. Nowak et al found that no bony contact and comminution were predictive of poorer results, whereas the degree of shortening was not a reliable predictor of sequelae. The authors utilized AP and 45° tilted views to characterize the fractures. Utilizing chest radiographs to compare the length of the injured clavicle with the noninjured side, Rasmussen et al reported that shortening of greater than 2 cm was not associated with poorer outcomes. Conversely, analyzing unilateral injured clavicle radiographs, Hill et al discovered that initial shortening at the fracture site of greater than or equal to 2 cm had a highly significant association with nonunion and the chance of an unsatisfactory result. Similarly, utilizing unilateral clavicle views and chest radiographs to determine the amount of initial shortening, Thomodsgard et al found that greater than 2 cm of initial shortening was associated with poorer outcomes. Other studies have shown that shortening as little as 15 mm on initial radiographs may be associated with worse functional results. De Giorgi et al suggested that the amount of shortening expressed as a percentage of the length of the native clavicle (based on the length of the contralateral clavicle on a chest radiograph) is more accurate than just the amount of shortening in millimeters because of the variability in the length of the clavicle in humans, and the authors reported that 9.7% shortening of the clavicle was associated with poorer results.

With the emphasis on the degree of shortening on plain radiographs as an important factor in functional outcomes and determining treatment, it is not surprising that there is much variability in the findings of the above studies. In our investigation, we found that there was only weak to no interobserver agreement and minimal intraobserver agreement on the amount of shortening when viewing AP and 30° cephalad views of the clavicle, despite having access to a calibrated measuring tool. The interobserver kappa values were particularly low in the ranges of shortening around the critical shortening values of 1 to 2 cm, which is the most important range when determining the outcomes of nonoperatively treated fractures. This poor reliability is reflective of the difficulty in determining the amount of shortening on plain unilateral radiographs of clavicle fractures, which tend to be oblique in nature with comminution. As a result, it is challenging to determine where the actual ends of the proximal and distal fragments are located. The poor reliability in determining the degree of shortening on the unilateral clavicle views may have contributed to the weak interobserver findings on whether operative treatment should be selected, since a majority of the evaluators in this investigation ranked the amount of shortening as the most variable in determining treatment. Similarly, the minimal intraobserver agreement on the amount of shortening could help explain why there was only moderate intraobserver agreement on the treatment.

However, we found moderate to strong interobserver and intraobserver agreement for both displacement (κ = 0.63 and 0.76, respectively) and comminution (κ = 0.75 and 0.80, respectively) when using standard plain radiographs. This is reflected in the Nowak et al study in which only the degree of displacement and the presence of comminution were predictive of outcome in nonoperatively treated midshaft clavicle fractures and in which shortening was not a reliable predictor for sequelae. As the authors discussed, displacement and comminution could be more reliably determined than shortening.

In our study, the following variables predicted whether surgery was recommended (P < .001): (1) odds of surgery were 2.26 if comminution was noted, holding displacement and the interaction between displacement and shortening constant; (2) the odds of surgery were 3.37 if there was displacement greater than 100% compared with displacement of 0%-49%, holding comminution and shortening constant. Therefore, the degree of displacement and the presence of comminution are more reliable than the degree of shortening in determining the need for surgery among orthopaedic surgeons because they are more reliably determined and measured radiographically.

Other investigations have also reported poor reliability in determining the degree of shortening and the need for surgery based on standard AP and cephalad views of the clavicle. Smekal et al compared posterior-anterior thorax radiographs to AP 15° cephalad panoramic views of the shoulder girdle in patients with healed clavicle fractures to determine which technique was most accurate in measuring shortening. They used the computed tomography scan as the gold standard and found that the posterior-anterior thorax radiograph was more accurate in measuring the degree of shortening than the cephalad panoramic view. Sharr et al found that a 15° caudal clavicle radiograph more reliably assessed shortening than a 15° cephalad view. The authors theorized that because the caudal view decreased the distance from the clavicle to the film, magnification was minimized, which decreased the chance for error in taking the radiograph. Finally, Austin et al investigated intra- and interobserver reliability in determining the need for surgery for clavicle fractures, using

---

References 2, 5, 6, 8, 10, 12, 13, 17-22, 26.
standard AP and 20° cephalic tilt views versus these 2 views and 45° cephalic tilt and 45° caudal tilt views. The authors reported that the addition of the last 2 views to the standard trauma views improved the intra- and interobserver reliability in determining the need for surgery, and the surgeons were more likely to treat the fractures operatively with the 2 additional views. The authors suggested that the 2 additional views improved visualization of the AP displacement.

As a result of our findings, we recommend that the degree of shortening on AP and 30° cephalad views not be used as the sole determining factor in deciding on surgery, owing to the poor agreement in measuring the amount of shortening among fellowship-trained shoulder/ sports medicine orthopaedic surgeons, despite having access to a electronically calibrated measuring tool. If the amount of shortening is utilized as the sole determinant of treatment, surgeons should consider obtaining other radiographic views, which have been shown in other investigations to more reliably determine the degree of shortening, including a posterior-anterior chest radiograph to determine the length of the contralateral clavicle or caudal views at varying angles. The main drawback to this, obviously, is more radiation exposure for the patient. It is important, though, to more reliably assess these factors to avoid under- or overestimating the severity of a clavicle fracture. This prevents undertreatment of the injury non-operatively, with the consequence of potential long-term sequelae from malunion, or overtreatment with surgery, which has potential increased risks of complications.7,9,29

Another option would be to utilize comminution or the degree of displacement on standard unilateral plain radiographs as the main determinants of treatment, since there is better agreement in determining these factors among fellowship-trained shoulder and sports medicine surgeons. This option needs to be investigated in future studies. With the weak interobserver agreement and only moderate intraobserver agreement in deciding on treatment when viewing AP and 30° cephalad unilateral clavicle radiographs, we need to develop better criteria or more reliable imaging to determine treatment options.

Study Limitations

There are a few limitations to our study. First of all, this was only an intra- and interobserver reliability study, not an accuracy study. We did not have a gold standard to which to compare our findings. Our main goal, however, was to determine whether fellowship-trained orthopaedic surgeons could agree on radiology findings and subsequent treatment. Second, we did not compare the reliability of our standard trauma views to other views to see if there was any improvement in the agreement with the additional views. Also, our radiographic views were not true orthogonal views. The radiographs that we included are representative of the radiographs typically obtained in a trauma setting or office setting. In addition, the goal of our study was to determine the intra- and interobserver agreement on measuring the amount shortening and degree of displacement, determining the presence of comminution, and determining treatment on standard plain radiographs of the clavicle, which are typically available when the patient presents for evaluation. Finally, we did not determine the exact reason why an evaluator in the study suggested surgical intervention versus nonsurgical intervention for each fracture. Our assumption was that each reviewer would utilize a combination of radiographic features in deciding on the treatment. We did, however, survey the surgeons after their evaluations to determine which variables were important, overall, in their determination of treatment.

CONCLUSION

Our study demonstrated moderate to strong inter- and intraobserver agreement for both displacement and comminution when utilizing unilateral AP and 30° cephalad plain radiographs. However, there was only weak to no interobserver agreement and minimal intraobserver agreement on the amount of shortening, despite observers having access to a calibrated measuring tool. Also, there was minimal interobserver and only moderate intraobserver agreement on whether operative treatment should be selected. Therefore, standard AP and 30° cephalad views are not reliable in determining the amount of shortening of clavicle fractures and are not reliable for determining the need for surgical intervention among fellowship-trained orthopaedic surgeons.

CONTRIBUTING AUTHORS

Keith Baumgartner, MD (Orthopedic Institute, Sioux Falls, South Dakota); John E. Kuhn, MD, Charles Cox, MD (Vanderbilt University, Nashville, Tennessee); Brian Wolf, MD, Matthew Bollier, MD (University of Iowa, Iowa City, Iowa); James L. Carey, MD, MPH (University of Pennsylvania, Philadelphia, Pennsylvania); Eric C. McCarty, MD, Armando F. Vidal, MD (University of Colorado, Boulder, Colorado); Edwin E. Spencer, MD (Knoxville Orthopedic Clinic, Knoxville, Tennessee); Brian Feeley, MD (UCSF Medical Center, San Francisco, California); Robert G. Marx, MD, MSc (Hospital for Special Surgery, New York, New York); Bruce Miller, MD, MS (University of Michigan, Ann Arbor, Michigan); Rick W. Wright, MD, Robert H. Brophy, MD, Matthew V. Smith, MD (Washington University, St Louis, Missouri); and Joseph Abboud, MD (Rothman Institute, Philadelphia, Pennsylvania).

ACKNOWLEDGMENT

The authors thank Margaret Knisley at The Ohio State University for her help in the literature search and obtaining the referenced papers for this manuscript.

REFERENCES
