Intra- and Interobserver Agreement in the Classification and Treatment of Distal Third Clavicle Fractures

Julie Y. Bishop,*† MD, Grant L. Jones,† MD, Brian Lewis,† MD, Angela Pedroza,† MPH, and MOON Shoulder Group‡
Investigation performed at OSU Sports Medicine, The Ohio State University, Columbus, Ohio, USA

Background: In treatment of distal third clavicle fractures, the Neer classification system, based on the location of the fracture in relation to the coracoclavicular ligaments, has traditionally been used to determine fracture pattern stability.

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

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

Methods: Thirty radiographs of distal clavicle fractures were randomly selected from patients treated for distal clavicle fractures between 2006 and 2011. The radiographs were distributed to 22 shoulder/sports medicine fellowship-trained orthopaedic surgeons. Fourteen surgeons responded and took part in the study. The evaluators were asked to measure the size of the distal fragment, classify the fracture pattern as stable or unstable, assign the Neer classification, and recommend operative versus nonoperative treatment. The radiographs were reordered and redistributed 3 months later. Inter- and intrarater agreement was determined for the distal fragment size, stability of the fracture, Neer classification, and decision to operate. Single variable logistic regression was performed to determine what factors could most accurately predict the decision for surgery.

Results: Interrater agreement was fair for distal fragment size, moderate for stability, slight for type IIB and III fractures, and moderate for treatment approach. Intrarater agreement was moderate for distal fragment size categories (κ = 0.50, P < .001) and Neer classification (κ = 0.42, P < .001) and substantial for stable fracture (κ = 0.65, P < .001) and decision to operate (κ = 0.65, P < .001). Fracture stability was the best predictor of treatment, with 89% accuracy (P < .001).

Conclusion: Fracture stability determination and the decision to operate had the highest interobserver agreement. Fracture stability was the key determinant of treatment, rather than the Neer classification system or the size of the distal fragment.

Keywords: distal clavicle fracture; Neer classification; fragment size; interrater agreement; intrarater agreement

Clavicle fractures account for up to 2.6% of all fractures, with 12% to 15% of these occurring in the distal clavicle.8 Although distal clavicle fractures are relatively uncommon, they are encountered more frequently by shoulder specialists and represent a challenging injury with high complication rates. Traditionally, nonoperative management has yielded successful results in stable fracture patterns with minimal displacement.13 In addition, some studies have shown that older and more sedentary patients have an acceptable clinical outcome even when fracture displacement is present and nonunion occurs.27 Nonoperative management of unstable fracture patterns in young, active, healthy patients has yielded poor results with regard to pain levels and function, often due to the increased rate of nonunion.4,19-22,28 Surgical treatment is more often performed for these patients. Given these findings, patient factors such as age, activity level, and associated comorbidities play an integral role in surgical decision making.

The Neer classification system has traditionally been used to determine fracture pattern stability and is based on the location of the fracture in relation to the coracoclavicular ligaments.3,20,21 Since determination of fracture pattern stability is a driving factor for surgical decision making, the reliability of this system is paramount. However, this system has not been evaluated in the literature to our knowledge. Robinson25 proposed a different classification system of all clavicle fractures in 1998. In this study,
distal clavicle fractures were classified as type III, with different subgroups based on displacement of the major fragments. Despite reporting a substantial interobserver reliability, this system has rarely been used in the literature, and thus, the Neer classification continues to be the gold standard for assessing these fractures.

When surgery is deemed necessary, the method of fixation is typically a subjective choice based on the preference of the treating surgeon, since no high level of evidence exists in the literature. Although the current literature describes many procedures for the treatment of these unstable fractures, there has been no consensus or gold standard determined, and many of the treatment options have high complication rates. Currently, the most widely used options for surgical fixation of these fractures are (1) a soft tissue reconstruction/repair with heavy suture, (2) internal fixation with a plate and screw construct, or (3) a Hook plate construct. However, these treatment options are still fraught with complications, and again the decision for treatment choice is very subjective. Often the size and quality of the distal clavicle fragment will guide the treatment options; however, it is unclear how accurately surgeons assess this size.

The purpose of this study was to improve our understanding of how accurately these fractures are assessed and which factors contribute to clinical decision making regarding fracture management. To accomplish this, we assessed the inter- and intraobserver reliability among a group of shoulder-trained orthopaedic surgeons evaluating distal clavicle fractures on plain radiographs to determine (1) the size of the distal clavicle fragment, (2) the stability of the fracture, (3) the Neer classification of the fracture, and (4) preference for operative or nonoperative management.

MATERIALS AND METHODS

This study was approved by the institutional review board of The Ohio State University. Radiographs of distal clavicle fractures were identified by searching the medical records of the principal investigators (J.Y.B. and G.L.J., both shoulder surgeons at The Ohio State University) using the International Classification of Diseases, Ninth Revision code 810.0 (clavicle fracture) and searching the time period from 2006 to 2011. Only radiographs of patients older than 18 years and younger than 89 years were retrieved. The medical record numbers were identified, and the picture archiving and communication system (PACS) archives at our institution (RADWeb) were reviewed. Radiographs were used for the study only if the initial injury films were present (anteroposterior and 30-degree cephalic tilt) and were of good radiographic quality to allow measurements to be taken. A power analysis was performed to determine the number needed to show statistical significance based on 10 raters. Using an α value of 0.05 and a power of 0.9, we arrived at 30 radiographs. Our data search retrieved 67 total patients with distal clavicle fractures. The radiographs were all reviewed, and 30 different radiographs were selected in an effort to select injuries across the entire spectrum of distal clavicle injury severity, with an even spread of ages, distal fragment sizes, and varying degrees of displacement. The radiographs were then de-identified.

The radiographs were transferred into a Microsoft PowerPoint file and resized based on the differences in image magnification to standardize measurements. The PowerPoint file was then converted into a PDF file in an effort to use the calibrated measuring tool embedded within the Adobe Acrobat Reader, which was used to determine the size of the distal fragment. To determine the accuracy of the conversion process into the PDF file, each image was measured in the PDF file using this system and then compared with the measurements obtained within the PACS system.

The PDF file containing the radiographs and detailed instructions on how to use the measuring tool within Adobe Acrobat Reader and the surveys/questionnaires were sent to the orthopaedic surgeons in a multicenter shoulder group that studies outcomes from shoulder surgeries (Multicenter Orthopaedics Outcomes Network [MOON] Shoulder Group). No demographic information was included in this packet. The senior authors (J.Y.B., G.L.J.) were excluded from participation, which left 22 members who were sent the initial survey. The members were given 2 months to complete the survey. Three months after the receipt of the final first round survey, the radiographs were scrambled into a different order and the survey and PDF files resent to the shoulder group members who participated in the first round. Fourteen surgeons completed each round of surveys.

In the questionnaires, the surgeons were asked to complete 4 tasks for each set of radiographs. First, they were asked to measure and record the size of the distal fragment in millimeters to 1 decimal point. Second, they assessed whether the fracture pattern was stable or unstable. Then they determined the type of fracture based on the Neer classification (I, IIA, IIB, or III). Finally, they recorded their recommendation for surgical or nonsurgical management. No specific instructions or criteria were given to the evaluating surgeons on how to determine the need for surgery, and all surgeons were instructed to use the criteria they would use in their own clinical practice. A copy of the Neer classification of distal clavicle fractures was also provided as a reference to all surgeons.

Statistical Methods

A modified k test was used to measure inter- and intraobserver reliability. Inter- and intrarater agreement was determined for the distal fragment size, stability of the fracture, Neer classification, and decision to operate by evaluating the k values, with k ≤ 0 = poor agreement, 0.01 to 0.2 = 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 1.0 = almost perfect agreement.

Distal fragment size was broken into 6 categories: less than 10.0 mm, 10.1 to 20.0 mm, 20.1 to 30.0 mm, 30.1 to 40.0 mm, 40.1 to 50.0 mm, and greater than or equal to 50.1 mm.
A fracture was defined as either stable or unstable. Neer classification was defined as I, IIA, IIB, or III. Evaluators were asked, “Would you treat this fracture operatively?” to which they were limited to yes/no. Single-variable logistic regression was performed to determine if fracture size, stability of the fracture, or Neer classification could accurately predict surgical intervention. Multivariable logistic regression was then tested with both fracture size and stability. Because fracture stability was included as part of the Neer classification, a multivariable logistic regression including both variables was not tested.

### RESULTS

#### Interrater Agreement

Interrater agreement is summarized in Table 1. Interrater agreement for increased distal fragment size was substantial for fragments $>50$ mm and moderate for fragments 30 to 40 mm, with only fair and slight agreement for other sizes. Stability demonstrated moderate interrater agreement, while the agreement for the Neer classification was fair overall. There was only slight agreement, however, on Neer IIB fractures. There was moderate agreement on indications for surgical treatment.

#### Intrarater Agreement

Intrarater agreement was moderate for fracture size categories ($\kappa = 0.50$, $P < .001$) and the Neer classification ($\kappa = 0.42$, $P < .001$). Intrarater agreement was substantial for determination of fracture stability ($\kappa = 0.65$, $P < .001$) and the decision to operate ($\kappa = 0.65$, $P < .001$).

#### Multivariable Logistic Regression

The odds of a stable fracture being recommended for surgery were less than 1% compared with unstable fractures ($P < .001$). In essence, $<1\%$ of fractures deemed stable were actually recommended for surgical treatment, whereas $89\%$ of those deemed unstable were recommended for surgery. Therefore, the decision to recommend surgery was based on whether the fracture pattern was determined to be stable. Thus, of 30 sets of radiographs, an average of $32\%$ were determined unstable by the group and hence recommended for surgery. Figure 1 presents examples of fracture patterns deemed stable and unstable by our reviewers. Adding in the size of the distal clavicle fracture fragment or the Neer classification rating did not significantly improve the prediction/recommendation of surgery for the distal clavicle fracture over just looking at stability alone ($P = .23$). The receiver operating characteristic (ROC) curve area for the model containing stability was significantly higher than the model containing classification and fracture fragment size. Knowing fracture pattern stability accurately predicted the decision to operate $89\%$ of the time (Figure 2).

### DISCUSSION

Distal clavicle fractures have always been notoriously difficult to treat. Thus, there has been debate as to not only the proper treatment technique but on identifying appropriate candidates for surgery. Most surgeons rely on the Neer classification to aid in decision making for surgery, and if a patient is deemed a candidate for surgery, many rely on
the size of the distal fragment to aid in choosing appropriate fixation.\textsuperscript{11,30} When evaluating how well experienced shoulder surgeons determine these important factors, our results have shown that the interrater agreement was fair for Neer classification, fair for distal fragment size, moderate for stability, and moderate for treatment approach. Classically, minimally displaced fractures of the distal clavicle have been treated nonoperatively, with good results reported in the literature. The chance of a nonunion is low, reported at only 5%, and can often be tolerated well, especially in older and more sedentary patients.\textsuperscript{22} However, unstable distal clavicle fractures have traditionally been recognized as more problematic and are often considered the one category of clavicle fracture that should be addressed operatively. Dating back to the work of Neer and other earlier reports, 22\textsuperscript{nd} to 33\textsuperscript{rd} of Neer type II distal clavicle fractures fail to heal with nonoperative treatment, and an additional 45\% to 67\% require more than the traditional 3 months for the fracture to heal.\textsuperscript{4,10,15-17,19-22,26,28} However, in older patients with more osteoporotic bone, there is a concern that surgical intervention can lead to operative complications. Several authors have reported acceptable results with nonoperative treatment in this age group.\textsuperscript{27} Some authors also consider this injury more in line with a type III acromioclavicular joint separation, and these injuries have proven to have acceptable results with nonoperative management. If nonoperative management fails to yield acceptable results, a reconstructive procedure can be performed at a distant time, similar to reconstructive options for the failed type III acromioclavicular joint separation.\textsuperscript{18,22,29}

Due to the concern for nonunion, Neer recommended stabilization of the unstable distal clavicle fracture.\textsuperscript{20} In his classification system, type I and III fractures are considered more stable, and the type II fracture, further subtyped into IIA and IIB, has traditionally been deemed unstable. The classification is based on the location of the coracoclavicular ligaments relative to the fracture site. Thus, in type II fractures, the medial fragment is disconnected from the remaining shoulder girdle and hence creates an unstable fracture pattern. Muscle forces elevate the medial fragment while the weight of the arm pulls the smaller lateral fragment downward through its connection with the coracoclavicular ligaments, possibly creating a large gap between the fracture fragments and thus predisposing to nonunion.

The Neer classification has been universally used for many years when determining fracture pattern stability. As noted above, unstable fracture patterns are believed to lead to a higher risk for nonunion and thus are treated operatively most often. However, it is unclear how well this classification system reproducibly determines fracture pattern stability. It has been noted in the literature that although there is a need to reconsider how we classify these fractures and determine treatment, there is a low overall rate of these injuries, which will make this very hard to accomplish.\textsuperscript{24} Unfortunately, data regarding evaluation and treatment of distal clavicle fractures consist primarily of retrospective studies or case series, many of which evaluate only 1 specific type of fixation method or technique. Prospective, randomized studies are needed to evaluate not only operative versus nonoperative care but also different treatment options when surgery is undertaken.

Our evaluation of the Neer classification showed an overall fair agreement between observers. However, the most important determining factor for surgery has been thought to be the type IIB Neer pattern. This pattern showed only slight agreement ($\kappa = 0.12$) between our shoulder-trained observers. This does call into question the sole use of the Neer classification system for determination of fracture pattern stability and the need for surgery. However, when simply asked if the fracture pattern was stable or unstable, agreement was much more consistent, with moderate agreement ($\kappa = 0.57$) between all observers. Furthermore, we found that the determination for surgery was almost exclusively decided by fracture pattern stability. Eighty-nine percent of the time, the decision for surgery was solely influenced by whether the fracture was deemed stable. Adding in the additional factors of the size of the fragment or the determination of the Neer classification system did not change or really influence this decision. Thus, perhaps just a more simple system is needed: stable or unstable. Of course, more research is needed to decide what truly is unstable versus stable. But it appears that this decision is much easier to agree upon among surgeons.

The size of this smaller lateral fragment is of utmost importance when surgery is deemed necessary as it will often guide the treating surgeon’s fixation options. As many different types of fixation now exist,\textsuperscript{17} this fragment must be large enough to accommodate the chosen fixation. Therefore, we believed that it was very important to assess how well we as surgeons can accurately determine the size of this fragment. Often, the most minimal size of the distal cluster of screws on a distal clavicle locking plate will be approximately a centimeter in length. Thus, optimal fixation would be with a fragment at least 1 cm or longer. Therefore, the assessment of this size will help decide whether the distal fragment of bone will accommodate a plate. One could surmise then that accurate and
consistent determination of the fragments that are 1 to 2 cm in length would be most important. Our interrater results showed overall fair agreement on fragment size. The most consistent assessment of fragment size was with the largest fragment (> 50.1 mm = substantial agreement). Unfortunately, distal fragments this size are not the most challenging fracture patterns when determining overall treatment plans. We found the 10.0- to 20-mm range the most problematic, with only slight and fair agreement, respectively. Our study used the standard views that are representative of the radiographs typically obtained in a trauma or office setting. Perhaps there would be an improvement in agreement on size of the distal fragment with the use of additional views, and this may be an area for further research. This is deemed important because inaccurate assessment of the true size of the distal clavicle could lead to inappropriate treatment options that would lead to a higher risk of failure of fixation.

We found that surgeons were very consistent with themselves, and the intraobserver agreement was moderate or substantial for distal fragment size, Neer classification, fracture pattern stability, and decision for surgery. But clearly, our interrater responses for the Neer system showed that they were not consistent with each other, and what one surgeon deems an unstable IIB, another may not. Thus, a classification system that is only internally consistent may not yield the most universally consistent results when used to determine the need for surgery. There were a few limitations to our study. First, this was only an intra- and interobserver reliability study and not an accuracy study. We did not have a gold standard to which to compare our findings. However, our main goal was to determine whether fellowship-trained shoulder surgeons could agree on key radiological findings and accepted current classification systems. Second, all decisions, especially whether to operate, were based solely on the radiographic appearance of the clavicle fractures. We did not take into account any patient factors, which certainly would influence decisions if included. We are aware that much more goes into operative decision making, such as patient age, activity level, smoking history, and so on. Thus, the results of our study should not be misinterpreted that one should operate on all unstable fracture patterns.

CONCLUSION

Our study has shown that the assessed stability of a distal clavicle fracture predicted the decision to operate 89% of the time. Therefore, as the intra- and interrater agreement was highest for fragment stability versus distal fragment size and the Neer classification, determination of fracture pattern stability may be a more appropriate and reliable method to classify distal clavicle fractures.

CONTRIBUTING AUTHORS

Keith Baumgarten, Orthopedic Institute, Sioux Falls, South Dakota; Jed Kuhn, Warren Dunn, and Charlie Cox, Vanderbilt University, Nashville, Tennessee; Brian Wolf, Carolyn Hettrich, and Matt Bollier, University of Iowa, Iowa City, Iowa; James Carey, John Kelly, and Brian Sennett, University of Pennsylvania, Philadelphia, Pennsylvania; Eric McCarty, Armando Vidal, Jonathan Bravman, and Sourav Poddar, University of Colorado, Boulder, Colorado; Edwin Spencer and Brian Holloway, Knoxville Orthopedic Clinic, Knoxville, Tennessee; Ben Ma, Christina Allen, and Brian Feeley, UCSF Medical Center, San Francisco, California; Robert Marx, Hospital for Special Surgery, New York, New York; Bruce Miller and Jim Carpenter, University of Michigan, Ann Arbor, Michigan; Rick Wright, Robert Brophy, and Matt Smith, Washington University, St Louis, Missouri; and Joseph Abboud, Rothman Institute, Philadelphia, Pennsylvania.

REFERENCES


For reprints and permission queries, please visit SAGE’s Web site at http://www.sagepub.com/journalsPermissions.nav