

Original Research

Pediatric Femoral Shaft Fracture Classification: An Intraobserver and Interobserver Reliability Study

Padam Kumar, BS¹; Richard A Hillesheim, MD²; Jeffrey R. Sawyer, MD²; James H. Beaty, MD²; David D. Spence, MD²; William C. Warner Jr., MD²; Benjamin W. Sheffer, MD²; Derek M. Kelly, MD²

¹University of Tennessee Health Science Center, College of Medicine, Memphis, TN; ²Department of Orthopaedic Surgery and Biomedical Engineering, University of Tennessee Health Science Center-Campbell Clinic, Memphis, TN

Correspondence: Derek M. Kelly, MD, Department of Orthopaedic Surgery and Biomedical Engineering, Le Bonheur Children's Hospital, University of Tennessee Health Science Center-Campbell Clinic, 1211 Union Ave., Suite 510, Memphis, TN 38104. E-mail: dkelly@campbellclinic.com

Received: February 7, 2022; Accepted: March 19, 2022; Published: May 1, 2022

DOI: 10.55275/JPOSNA-2022-0036

Abstract:

Purpose: Fracture stability is important in choosing the optimal treatment for pediatric femoral fractures, although there is no consensus for characterizing a fracture as “stable” or “unstable.” The authors sought to measure interobserver and intraobserver reliability in classifying femoral fracture stability and examined the relationship between fracture ratio and perceived fracture stability and morphology.

Methods: Fracture ratios were calculated from anteroposterior and lateral radiographs from 65 children aged 5 to 12 years, who were treated for femoral shaft fractures at a level 1 pediatric trauma center. Deidentified radiographs were placed into a PowerPoint presentation in random order and were shown to six fellowship-trained pediatric orthopaedic surgeons at two time points 4 months apart. Raters classified stability as “stable/unstable” and morphology as “spiral/oblique/transverse.” Cohen and Fleiss kappa (k) values were calculated to determine intraobserver and interobserver reliability. Generalized linear modeling was used to compare FR to rater fracture stability and morphology.

Results: The mean k for fracture stability for all raters was 0.68 (strong intraobserver agreement). The k for fracture stability during round 1 was 0.53 (67.7% moderate interobserver agreement). The k for fracture stability during round 2 was 0.68 (75.4% strong interobserver agreement). The mean k for fracture morphology for all raters was 0.79 (strong intraobserver agreement). The k for fracture morphology during round 1 was 0.38 (15.4% fair agreement). The k for fracture morphology during round 2 was 0.46 (24.6% moderate agreement). The average anteroposterior fracture ratio in fractures deemed stable was 1.32 compared with 1.78 in unstable fractures ($P < 0.001$). The average lateral fracture

ratio in stable fractures was 1.34 compared with 2.10 in unstable fractures ($P < 0.001$). Average anteroposterior and lateral fracture ratios were highest in spiral fractures and lowest in transverse fractures ($P < 0.003$).

Conclusions: Raters demonstrated strong intraobserver and interobserver agreement in classifying radiographic femoral fracture stability. Anteroposterior and lateral fracture ratios were significantly higher in unstable fractures.

Level of Evidence: Level III

Key Concepts:

- This study sought to measure intraobserver and interobserver reliability in determining pediatric femoral fracture morphology and stability.
- The authors also were interested in seeing if an objective measurement, the fracture ratio (FR), could accurately predict fracture morphology and stability.
- A strong intraobserver agreement in defining both fracture stability and morphology was found.

Introduction

Orthopaedic injuries are one of the most common causes of pediatric inpatient hospitalization. A 2014 study found that among pediatric orthopaedic injuries requiring hospitalization, femoral fractures are the most common reasons for admission and account for the highest monetary cost of hospitalization.¹

Current guidelines from the American Association of Orthopaedic Surgeons (AAOS) utilize an age-based approach for treatment of femoral shaft fractures.² Recommendations include: Pavlik harness or spica casting for infants age 6 months or younger, early spica casting or traction with delayed spica casting for children age 6 months to 5 years old, flexible elastic intramedullary nailing (EIN) for children age 5 to 11 years old, and rigid trochanteric nailing, submuscular plating, or flexible intramedullary nailing for skeletally immature children age 11 years or older. Of these, only EIN for children 5 to 11 years old carries a “strong” recommendation.² However, studies have shown that there is considerable variability and deviation from these AAOS guidelines in the treatment of pediatric femoral shaft fractures.^{3,4}

This difference exists because age is not the only factor influencing treatment choice. Other important

factors include geographic bias (e.g., regional care norms), patient weight, and fracture morphology.^{5,6} Determination of fracture morphology is subjective and surgeon dependent. A 2014 study found that classifying pediatric femoral shaft fracture morphology was highly variable among pediatric orthopaedic surgeons, emergency room physicians, and musculoskeletal radiologists.⁷ Notably, the orthopaedic surgeons in that study demonstrated strong intraobserver agreement and moderate interobserver agreement.

Length stability of a fracture is another characteristic often cited as a factor in determining treatment type. Length unstable fractures are often thought to be more complex and require more stable fixation than length stable fractures, though recent studies have called this notion into question.⁸⁻¹¹ Regardless, there is no consensus on objective criteria for what constitutes a fracture as stable or unstable, with some authors describing comminuted fractures as unstable and others listing spiral or long oblique fractures as unstable.^{8,12,13} Although a clinically relevant, agreed upon definition remains elusive, papers have recommended using length stability or fracture morphology as criterial for determining treatments.^{5,6,14,15}

The purpose of this study was to measure the intraobserver and interobserver reliability in classifying pediatric femoral shaft fracture stability and morphology among fellowship-trained pediatric orthopaedic surgeons. The authors also wanted to determine if fracture ratio (FR), an objective measurement, could be correlated to the observers' responses. This could allow for a more objective method of classifying fracture morphology and stability. The hypotheses were that both intraobserver and interobserver reliability would be strong and that FR would be correlated with observed fracture morphology and perceived fracture stability.

Materials and Methods

This study was approved by the institutional review board (IRB) at our institution. ICD-10 codes were obtained to identify children aged 5 to 12 years who were treated for a closed femoral shaft fracture at a pediatric level 1 trauma center from September 2015 to July 2019. An a priori power analysis determined that for six raters and an alpha of 0.05, a sample size of 65 patients was required to achieve optimal power. Exclusion criteria included patients with insufficient radiographs, fractures secondary to gunshots, fracture comminution, fractures in the proximal and distal fourth of the femur, and patients with metabolic bone disease. Fractures secondary to gunshot wounds and fractures with comminution were excluded because the authors intended to study fractures with only one distinct fracture line.

Initial injury anteroposterior and lateral femoral radiographs were obtained from the electronic medical record for each patient. Fracture length and bone diameter were measured on both anteroposterior and lateral radiographs. FR was defined as fracture length divided by bone diameter at the level of the fracture (Figure 1). Deidentified anteroposterior and lateral radiographs for each patient were placed in a PowerPoint (Microsoft, Redmond, WA) slideshow presentation in randomized order and were shown to six fellowship-trained pediatric orthopaedic surgeons from a single institution 4 months apart (round 1 and round 2). Rater numbers were randomly assigned. Slides were advanced by a preset 10 second

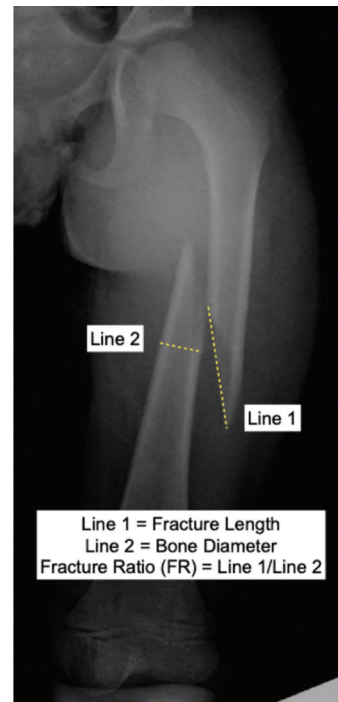


Figure 1. Fracture ratio (FR) was calculated by dividing the fracture length (Line 1) by the bone diameter of the distal fragment (Line 2).

timer. Reviewers were given an answer sheet and were told to circle “unstable” or “stable” for the fracture stability category and “spiral,” “oblique,” or “transverse” for the fracture morphology category. Because there are currently no well-defined definitions of fracture stability or fracture morphology in the literature, the reviewers were simply asked to rate the fractures without being provided specific definitions of either. Furthermore, the reviewers were only provided with the initial fracture films since images of the ultimate treatment could have influenced their decisions. The reviewers were aware of the purpose and hypothesis of the study.

All statistical analysis was conducted using R version 4.1.0. *P* values ≤ 0.05 were considered statistically significant. Intraobserver reliability for all observers was calculated using the Cohen kappa (*k*) statistic, and interobserver reliability was calculated using the Fleiss *k* statistic. The following cutoffs were used to interpret the reported *k* values: zero, no agreement; 0.10 to 0.2,

slight agreement; 0.21 to 0.4, fair agreement; 0.41 to 0.6, moderate agreement; 0.61 to 0.8, substantial agreement, 0.81 to 0.99, nearly perfect agreement; and 1, perfect agreement (<https://www.statology.org/cohens-kappa-statistic>, March 15, 2022). Percent agreement values ranged from 0% to 100%, with 0% referring to no agreement and 100% referring to complete agreement. To examine the relationship between FR and reviewer-reported fracture morphology and stability, generalized linear modeling was utilized. The models used round 1 reviewer observations for each patient and controlled for rater. Entries with null responses were excluded. Results were reported as mean ± standard error.

Results

Participant Characteristics

A total of 95 patients were initially evaluated based on search criteria. Thirty patients were excluded per the exclusion criteria outlined in the methods section. The average age of patients included in this study was 7.96 ± 1.77 years, with a range of 5 to 11.3 years.

Intraobserver Reliability - Stability

The mean k for fracture stability for all participants was 0.68 (substantial intraobserver agreement). The lowest k value was 0.36, corresponding to 75.4% agreement. The highest k value was 0.84, corresponding to a percent agreement of 96.9%. Rater 1 did not choose an answer for two radiographs, one in round 1 and the other in round 2 (Table 1).

Table 1. Intraobserver Reliability in Classifying Femoral Fracture Stability

| Rater | Cohen k | Estimated % agreement | Categorization of agreement | Mean k (categorization of agreement) |
|-------|---------|-----------------------|-----------------------------|--------------------------------------|
| 1 | 0.36 | 75.4% | Fair | 0.68 (substantial) |
| 2 | 0.63 | 86.2% | Moderate | |
| 3 | 0.72 | 89.2% | Substantial | |
| 4 | 0.72 | 92.3% | Substantial | |
| 5 | 0.80 | 92.3% | Substantial | |
| 6 | 0.84 | 96.9% | Nearly perfect | |

k, kappa

Interobserver Reliability - Stability

The k for fracture stability during round 1 was 0.53, indicating moderate interobserver agreement. This corresponded to 67.7% interobserver agreement. The k for fracture stability during round 2 was 0.68, indicating substantial interobserver agreement. This corresponded to 75.4% agreement (Table 2).

Intraobserver Reliability - Morphology

The mean k for fracture morphology for all participants was 0.79 (substantial agreement). The lowest k value was 0.67, corresponding to 73.8% agreement. The highest k value was 0.77, corresponding to a percent agreement of 89.2%. Rater 1 did not choose an answer for three radiographs, two in round 1 and one in round 2 (Table 3).

Interobserver Reliability - Morphology

The mean k for fracture morphology during round 1 was 0.38, indicating fair interobserver agreement. This corresponded to 15.4% interobserver agreement. The k for fracture morphology during round 2 was 0.46, indicating moderate interobserver agreement. This corresponded to 24.6% agreement (Table 2).

Fracture Ratio Calculations

Of the 390 total possible observations (six raters each responding to 65 patients), two were excluded because of null responses. The remaining 388 observations were utilized for the generalized linear models.

Table 2. Interobserver Reliability in Classifying Femoral Fracture Stability and Morphology

| | Round | Fleiss k | Categorization of agreement | Estimated % agreement |
|---------------------|-------|----------|-----------------------------|-----------------------|
| Fracture stability | 1 | 0.53 | Moderate | 67.7% |
| | 2 | 0.68 | Substantial | 75.4% |
| Fracture morphology | 1 | 0.38 | Fair | 15.4% |
| | 2 | 0.46 | Moderate | 24.6% |

k, kappa

Table 3. Intraobserver Reliability in Classifying Femoral Fracture Morphology

| Rater | Weighted Cohen k | Estimated % agreement | Categorization of agreement | Mean k (categorization of agreement) |
|-------|------------------|-----------------------|-----------------------------|--------------------------------------|
| 1 | 0.67 | 73.8% | Substantial | 0.68 (substantial) |
| 2 | 0.76 | 81.5% | Substantial | |
| 3 | 0.84 | 83.1% | Nearly perfect | |
| 4 | 0.75 | 73.8% | Substantial | |
| 5 | 0.81 | 78.5% | Nearly perfect | |
| 6 | 0.90 | 89.2% | Nearly perfect | |

k, kappa

The mean anteroposterior FR for unstable fractures was significantly higher (1.78 ± 0.062) compared with the mean anteroposterior FR for stable fractures (1.32 ± 0.029 ; $P < 0.0001$). The mean lateral FR for unstable fractures was significantly higher (2.10 ± 0.076) compared with the mean lateral FR for stable fractures (1.34 ± 0.035 ; $P < 0.0001$).

Using transverse fracture morphology as the reference category, mean anteroposterior FR were highest for spiral fractures (1.86 ± 0.061 ; $P < 0.0001$), followed by oblique fractures (1.39 ± 0.039 ; $P = 0.0022$), and transverse fractures (1.21 ± 0.041 ; $P < 0.0001$). Using transverse fracture morphology as the reference category, mean lateral FR were highest for spiral fractures (2.22 ± 0.074 ; $P < 0.0001$), followed by oblique fractures (1.42 ± 0.047 ; $P = 0.026$), and transverse fractures (1.21 ± 0.050 ; $P < 0.0001$) (Figure 2).

Discussion

There is limited consensus on the optimal treatment of pediatric femoral shaft fractures. Studies have shown that in many cases, surgeon and regional preferences prevail over national clinical practice guidelines.²⁻⁴ In addition to patient age, treatment decisions for pediatric femoral shaft fractures are influenced by several other factors, including the surgeon's determination of fracture morphology and interpretation of length stability.^{5,6,14,15}

This study sought to measure intraobserver and interobserver reliability in determining pediatric femoral fracture morphology and stability. The authors were also interested in seeing if an objective measurement (fracture ratio) was correlated with observed fracture morphology and perceived stability. A strong intraobserver agreement in defining both fracture stability and morphology was found. During round 1, there was moderate interobserver

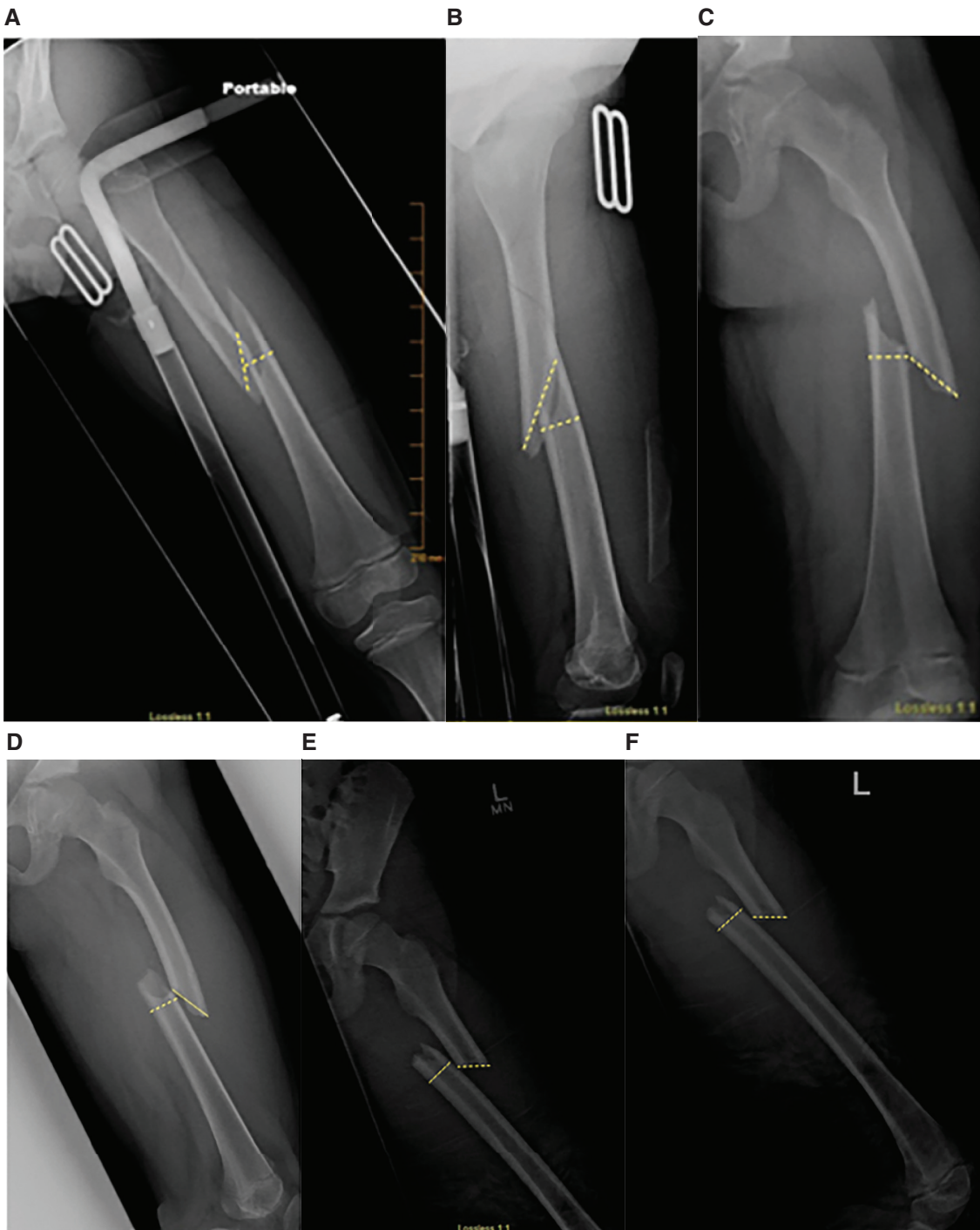


Figure 2. Prototypical morphology of spiral (A and B), oblique (C and D), and transverse fractures. Anteroposterior fracture ratios: spiral 1.86 (A), oblique 1.39 (C), and transverse 1.21 (E). Lateral fracture ratios: spiral 2.22 (B), oblique 1.86 (D), and transverse 1.21 (F).

agreement in defining fracture stability and fair interobserver agreement in defining fracture morphology. During round 2, there was strong interobserver

agreement in defining fracture stability and moderate interobserver agreement in defining fracture morphology. Statistical modeling demonstrated that unstable fractures

had significantly higher anteroposterior and lateral FR than stable fractures.

Intraobserver and interobserver reliability in classifying pediatric femoral shaft fractures was measured in a 2014 study by Thompson et al.⁷ There were 14 participants in their study, including seven pediatric orthopaedic surgeons, five pediatric emergency room physicians, and two fellowship-trained musculoskeletal radiologists. The authors found that among all participants, there was moderate interobserver agreement in categorizing fracture stability ($k = 0.48$). This was true for the orthopaedic surgeon subset as well, which demonstrated moderate interobserver agreement ($k = 0.54$). In the present study, there was fair interobserver agreement during round 1 ($k = 0.38$) and moderate interobserver agreement during round 2 ($k = 0.46$). The lower interobserver agreement observed during round 1 could have been due to one rater not choosing an answer regarding fracture morphology for two radiographs during that round. During statistical analysis, this was included as an additional response variable and led to a lower k value. Additionally, it is more challenging to get complete agreement when there are three options to choose from. Regarding intraobserver reliability, the Thompson et al.⁷ study found that the overall mean k value for all raters was 0.65 (strong intraobserver agreement). For the orthopaedic surgeon group, the mean k was 0.71 (strong intraobserver agreement). In the present study, the mean k for fracture morphology was 0.79 (strong intraobserver agreement). The similarities in intra- and interobserver agreement across both studies indicates that internally, pediatric orthopaedic surgeons have a strong and reproducible interpretation of femoral fracture morphology. However, there were a few key differences between the 2014 study and current study. Although both studies examined reliability in classifying femur fracture morphology, the current study also measured reliability in classifying femoral fracture stability and attempted to correlate both fracture morphology and fracture stability with FR. Additionally, the current study only utilized fellowship-trained, pediatric orthopaedic surgeon observers.

Within the pediatric orthopaedic literature, there are varying definitions of what constitutes an unstable femoral shaft fracture. In 2009, Kocher et al.¹² defined an unstable fracture as “comminuted or spiral fractures with greater than 2 cm of shortening.”¹² In a 2010 retrospective cohort study, Sink et al.¹³ defined an unstable fracture as “comminuted or long oblique (the length of the fracture is longer than the diameter of the femur at the level of the fracture).” An additional definition for an unstable fracture pattern was provided by Miller et al.⁸ Their retrospective review considered femoral shaft fractures unstable if they were “comminuted, spiral, or oblique fractures in which the fracture length was more than twice its width.” In the present study, reviewers were asked to select either “unstable” or “stable” after reviewing each fracture for ten seconds. By asking surgeons to decide quickly on stability, the authors hoped to capture their initial, unbiased response. Overall, there was strong intraobserver agreement among all orthopaedic surgeons. Percent intraobserver agreement ranged from 75.4% to 96.9%. These data suggest that although there is no consensus definition of fracture stability, each individual surgeon has developed a method of interpreting a fracture’s inherent stability. Interobserver agreement was moderate during round 1 and strong during round 2. These data suggest that generally, orthopaedic surgeons have an intuitive understanding of fracture stability and are significantly more likely to agree with each other than would be expected by chance alone.

To define fracture morphology and stability more objectively, FR was calculated for each radiograph. This value was calculated by dividing the fracture length by the bone diameter at the level of the fracture (Figure 1). To the authors’ knowledge, FR has been utilized in three studies to date. Thompson et al.⁷ found that at an anteroposterior FR less than 1.47, reviewers were likely to classify a fracture as transverse, and at an anteroposterior FR of 3.45 or more, reviewers were likely to classify a fracture as spiral. A 2015 study by Murphy et al.¹⁶ found that pediatric femoral fractures caused by nonaccidental trauma (NAT) had

significantly lower mean anteroposterior FR compared with those with confirmed accidental trauma.¹⁶ In that study, fractures classified as transverse were more commonly associated with NAT than spiral or oblique fractures. Vaughan et al.¹⁷ measured the effects of age and torsional rate on FR in a porcine femur model. They found that FR increased with specimen age and increased rate of torsional force applied. In the present study, anteroposterior and lateral FR were significantly higher in fractures deemed unstable compared with those deemed as stable. Additionally, the data show that spiral fractures had significantly higher anteroposterior and lateral FR compared with oblique and transverse fractures. Transverse fractures had the lowest anteroposterior and lateral FR.

There were limitations in this study. First, one rater did not complete responses for all radiographs. Specifically, this rater did not submit a response for fracture stability once in round 1 and another time in round 2. This rater also failed to complete a response for fracture morphology on three occasions: twice in round 1 and once in round 2. These missing responses were used in the statistical analysis and likely led to lower k values. Additionally, having only 10 seconds to determine the morphology and stability of a fracture without a picture archiving and communications system (PACS) may not be clinically relevant. In practice, a surgeon would likely sit near a computer monitor, optimize the radiograph for viewing, and take time to methodically scan the image. The surgeon would also have the benefit of examining the patient. Finally, the raters were not asked to classify fractures based on their planned treatment; rather, they were asked to rate the fracture stability and morphology on fracture characteristics alone.

To the authors' knowledge, this is the first study to measure interobserver and intraobserver reliability in classifying femoral fracture stability. Additionally, this study lends credence to the utility of FR as an objective tool in predicting pediatric femoral fracture stability and morphology. Now that fracture ratio has been defined and linked to the more subjective classifications of

fracture stability and fracture morphology, future studies comparing fixation methods and outcomes for FR would be useful.

Disclaimer

J. Beaty: Financial relationship with Elsevier Health; D. Kelly: Financial relationship with Elsevier Health; J. Sawyer: Financial relationship with DePuy (Johnson and Johnson, Co.), Elsevier Health, and OrthoPediatrics; D. Spence: Financial relationships with Elsevier Health and OrthoPediatrics; W. Warner: Non-financial relationship with Medtronic/Sofamor Danek and financial relationships with Elsevier Health and Wolters Kluwer Health-Lippincott Williams and Wilkins. The authors have no conflicts of interest to report.

References

1. Nakaniida A, Sakuraba K, Hurwitz EL. Pediatric orthopaedic injuries requiring hospitalization: epidemiology and economics. *J Orthop Trauma*. 2014;28(3):167-172.
2. American Academy of Orthopaedic Surgeons Board of Directors. [AAOS web site]. Treatment of pediatric diaphyseal femur fractures: evidence-based clinical practice guidelines. December 5, 2020. Available at: <https://www.aaos.org/globalassets/quality-and-practice-resources/pdf/pdfcpg.pdf>. Accessed November 1, 2021.
3. Oetgen ME, Blatz AM, Matthews A. Impact of clinical practice guideline on the treatment of pediatric femoral fractures in a pediatric hospital. *J Bone Joint Surg Am*. 2015;97(20):1641-1646.
4. Roaten JD, Kelly DM, Yellin JL, et al. Pediatric femoral shaft fractures: a multicenter review of the AAOS clinical practice guidelines before and after 2009. *J Pediatr Orthop*. 2019;39(8):394-399.
5. Hosalkar HS, Pandya NK, Cho RH, et al. Intramedullary nailing of pediatric femoral shaft fracture. *J Am Acad Orthop Surg*. 2011;19:472-481.
6. Flynn JM, Schwend RM. Management of pediatric femoral shaft fractures. *J Am Acad Orthop Surg*. 2004;12:347-359.
7. Thompson NB, Kelly DM, Warner WC Jr, et al. Intraobserver and interobserver reliability and the role of fracture morphology in classifying femoral shaft fractures in young children. *J Pediatr Orthop*. 2014;34(3):352-358.
8. Miller DJ, Kelly DM, Spence DD, et al. Locked intramedullary nailing in the treatment of femoral shaft fractures in children younger than 12 years of age: indications and preliminary report of outcomes. *J Pediatr Orthop*. 2012;32(8):777-780.
9. Kuremsky MA, Frick SL. Advances in the surgical management of pediatric femoral shaft fractures. *Curr Opin Pediatr*. 2007;19:51-57.
10. Siddiqui AA, Abousamra O, Compton E, et al. Titanium elastic nails are a safe and effective treatment for length unstable pediatric femur fractures. *J Pediatr Orthop*. 2020;40(7):e560-e565.
11. Mussell EA, Jardaly A, Gilbert SR. Length unstable femoral fractures: misnomer? *World J Orthop*. 2020;11(9):380-390.
12. Kocher MS, Sink EL, Blasler DR, et al. Treatment of pediatric diaphyseal femur fractures. *J Am Acad Orthop Surg*. 2009;17(11):718-725.
13. Sink EL, Faro F, Polousky J, et al. Decreased complications of pediatric femur fracture with a change in management. *J Pediatr Orthop*. 2010;30:633-637.

14. Narayanan UG, Hyman JE, Wainwright AM, et al. Complications of elastic stable intramedullary nail fixation of pediatric femoral fractures, and how to avoid them. *J Pediatr Orthop*. 2004;24(4):363-369.
15. Sink EL, Gralla J, Repine M. Complications of pediatric femur fractures treated with titanium elastic nails: a comparison of fracture types. *J Pediatr Orthop*. 2005;25(5):577-580.
16. Murphy R, Kelly DM, Moisan A, et al. Transverse fractures of the femoral shaft are a better predictor of nonaccidental trauma in young children than spiral fractures are. *J bone Joint Surg Am*. 2015;97(2):106-111.
17. Vaughan PE, Wei F, Haut RC. Effects of age and rate of twist on torsional fracture patterns in infant porcine femora. *J Clin Orthop Trauma*. 2020;11(2):281-285.