Intraobserver and Interobserver Reliability of Radiographic Analysis of Proximal Humerus Fractures in Adolescents

Michelle C. Burke, MS, Christopher Minnock, MD, Christopher B. Robbins, PhD, Matthew D. Abbott, MD, Michelle S. Caird, MD, Frances A. Farley, MD, Jacob Kirsch, MD, Jared Thomas, MD, and Ying Li, MD

Background: Multiple studies have shown low intrarater and interrater agreement of radiographic classification systems for proximal humerus fractures (PHFs) in adults. There is no standardized method of measuring angulation of pediatric PHFs, nor is there consensus as to the amount of angulation and displacement that require operative fixation of adolescent PHFs. We propose a new standardized method to measure fracture angulation that is similar to the method used to measure the epiphyseal-shaft angle for slipped capital femoral epiphysis. The primary purpose of this study was to evaluate the intraobserver and interobserver reliability of our proposed method compared with a nonstandardized method. The secondary purpose was to evaluate the intrarater and interrater agreement of the Neer and Horowitz (NH), and Salter-Harris (SH) classification systems.

Methods: Seven raters evaluated 26 deidentified anteroposterior shoulder radiographs of patients 10 to 16 years of age with PHFs. Raters classified each fracture using the NH and SH systems, and used their own method to measure fracture angulation. This process was repeated 2 weeks later. During the second round, raters also measured fracture angulation using our proposed standardized method. Two weeks after the second round, raters reevaluated the radiographs using the standardized method. Intraclass correlation coefficients were calculated.

Results: Excellent intraobserver and interobserver agreement was achieved for the standardized method of measuring fracture angulation. All of the raters had an intrarater reliability classified as excellent (>0.80) using the standardized method. Good intrarater and excellent interrater agreement was achieved when raters used their own fracture angulation measurement method but wide confidence intervals suggested that the results were less precise. Fair to moderate intrarater and interrater reliability was seen for the NH and SH classifications.

Conclusions: Our standardized method for measuring angulation in adolescent PHFs demonstrated excellent intrarater and interrater reliability. We propose that this technique may be a more precise method of measuring fracture angulation and this method should be used in future studies that evaluate indications for operative management of adolescent PHFs.

Level of Evidence: Level III—diagnostic.

Key Words: pediatric proximal humerus fracture, reliability, angulation

(*J Pediatr Orthop* 2017;00:000–000)

Pediatric proximal humerus fractures (PHFs) constitute approximately 2% of all pediatric fractures.[1,2] These fractures are usually the result of moderate energy trauma, such as from motor vehicle accidents or sports participation. The most commonly used classification systems for pediatric PHFs are the Neer and Horowitz (NH),[3] and Salter-Harris (SH)[4] classification systems. The great remodeling potential of the pediatric proximal humerus and significant shoulder range of motion in multiple planes allow most pediatric PHFs to be successfully treated with nonoperative management.[5] Mildly displaced fractures (NH grades I and II) can be treated nonoperatively with a sling or hanging arm cast. Adolescents with more displaced (NH grades III and IV) or angulated fractures may require surgical intervention secondary to decreased remodeling capacity. Operative indications for NH grades III and IV fractures in adolescents vary widely in the literature.[2,6–12]

Multiple studies have shown low intraobserver and interobserver reliability of radiographic classification systems for PHFs in adults.[13–24] There are no studies that have investigated the intraobserver and interobserver agreement of radiographic classification systems for pediatric PHFs. In addition, there is no standardized method of measuring angulation of pediatric PHFs. As there is no consensus with regard to management of displaced PHFs in adolescents, it is important to determine the reliability of the radiographic measures and classification systems that are commonly used to guide treatment.

We propose a new standardized method to measure angulation for adolescent PHFs that is similar to the method used to measure the epiphyseal-shaft angle for slipped capital femoral epiphysis[25] (Fig. 1). The epiphyseal-shaft angle is the most commonly used classification system to determine the severity of a slipped capital femoral

From the Department of Orthopaedic Surgery, C.S. Mott Children's Hospital, University of Michigan, Ann Arbor, MI.

None of the authors received financial support for this study.

The authors declare no conflicts of interest.

Reprints: Ying Li, MD, Department of Orthopaedic Surgery, C.S. Mott Children's Hospital, University of Michigan, 1540 E. Hospital Drive, SPC 4241, Ann Arbor, MI 48109-4241. E-mail: yingyuli@med.umich.edu.

DOI: 10.1097/BPO.0000000000001083

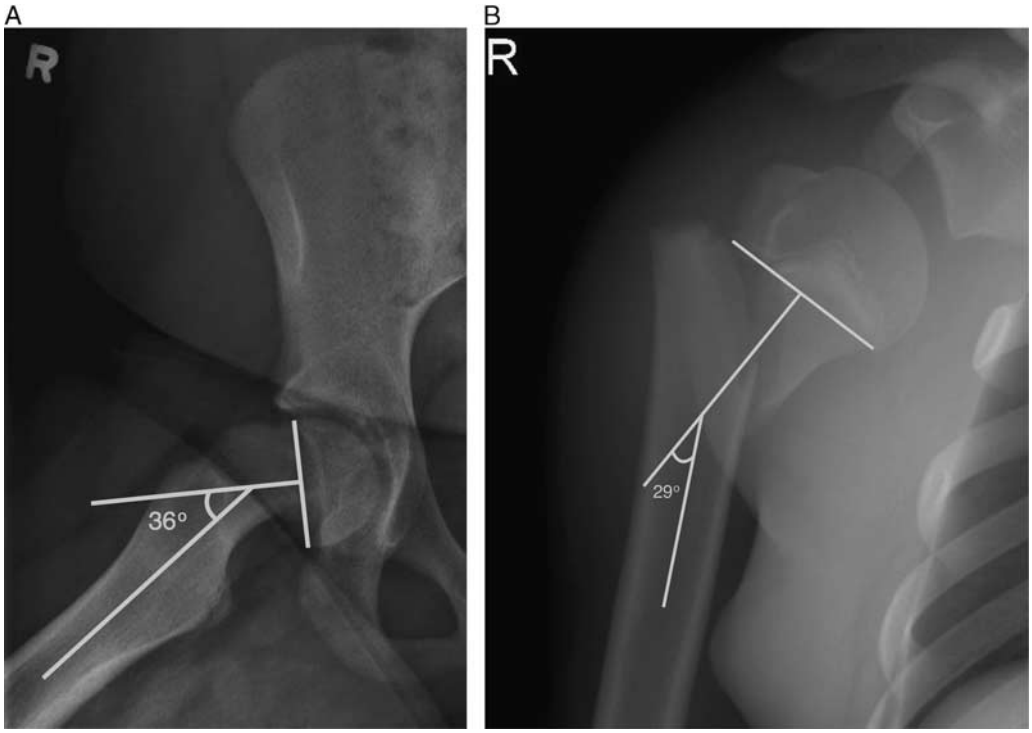

FIGURE 1. A, Right hip radiograph in a patient with slipped capital femoral epiphysis. The epiphyseal-shaft angle is measured on the frog-leg lateral pelvis radiograph. The angle is formed by a line perpendicular to the proximal femoral physis and the axis of the femoral shaft. B, Right shoulder radiograph in a patient with a proximal humerus physeal fracture. Our proposed measurement method for fracture angulation is similar to the measurement method for the epiphyseal-shaft angle in slipped capital femoral epiphysis. The angle is formed by a line perpendicular to the proximal humeral physis and the axis of the humeral shaft.

epiphysis. The angulation that results when pediatric PHFs displace through the physis can be measured using a similar method as with slipped capital femoral epiphysis.

The primary purpose of this study was to evaluate the intraobserver and interobserver reliability of this proposed method of measuring angulation of adolescent PHFs compared with a nonstandardized method. The secondary purpose of this study was to evaluate the intraobserver and interobserver agreement of the NH and SH classification systems for adolescent PHFs.

METHODS

Exemption status was granted by our Institutional Review Board before the start of this study. Seven raters with varying levels of experience participated in this study: 1 senior fellowship-trained pediatric orthopaedic surgeon, 1 mid-career fellowship-trained pediatric orthopaedic surgeon, 2 junior fellowship-trained pediatric orthopaedic surgeons, 1 pediatric orthopaedic surgery fellow, one-fifth year orthopaedic surgery resident, and one-third year orthopaedic surgery resident.

A power analysis demonstrated that based on a null intraclass coefficient (ICC) value of 0.6 and a hypothesized value of 0.8 with $P < 0.05$ and 95% confidence interval (CI), a minimum of 26 radiographs would need to be evaluated by each rater. Twenty-six anteroposterior shoulder radiographs were selected from patients 10 to 16 years of age who were treated for a PHF at our institution from January 2005 to January 2016. Exclusion criteria were patients with a closed proximal humeral physis and patients with poor quality radiographs. There were 20 males and 6 females, with an average age of 12.6 years. The radiographs were selected to include a wide range of patient ages, fracture displacement, fracture angulation, and fracture patterns. The radiographs were deidentified and were loaded onto CD-ROMS as Digital Imaging and Communication in Medicine files. Raters uploaded the files into the Digital Imaging and Communication in Medicine viewer, iQ-VIEW version 2.8.0 (IMAGE Information Systems Ltd, Charlotte, NC), to perform measurements. All raters received a document demonstrating the NH and SH classification systems.

For the first round, raters classified each fracture using the NH and SH classification systems, and used their own methodology to measure fracture angulation. Two weeks after the last rater returned their first round of measurements, raters were provided with a second disc with the same radiographs rearranged in a different order. Raters were also provided with a document demonstrating our proposed standardized method of measuring fracture angulation (the angle formed by a line perpendicular to the proximal humeral physis and the axis of the humeral shaft). Raters were asked to repeat the analysis that they had performed during the first round and then to measure fracture angulation using the standardized method. Two

weeks after the last rater submitted their second round of measurements, raters were asked to reanalyze fracture angulation on the radiographs from the first round of measurements using the standardized method.

Statistical analysis was performed using SPSS version 22.0 (IBM Corp, Armonk, NY). Intrarater and interrater agreement was determined by evaluating κ values for categorical data and ICCs for continuous data, according to Cohen[26] and Fleiss.[27] The ICCs were evaluated using a random effects model with absolute agreement. The interpretation scale described by Landis and Koch[28] was used: 0 to 0.20 = poor agreement, 0.20 to 0.40 = fair agreement, 0.40 to 0.60 = moderate agreement, 0.60 to 0.80 = good agreement, > 0.80 = excellent agreement.

RESULTS

Intrarater and interrater reliability results are shown in Table 1. Excellent intraobserver agreement was found with the standardized method of measuring fracture angulation, whereas good intraobserver agreement was seen when raters used their own measurement method. All raters achieved excellent intrarater reliability using the standardized method (Table 2). However, only the 4 pediatric orthopaedic surgery fellowship-trained surgeons had excellent intrarater reliability when raters used their own method.

Although excellent interobserver agreement was found with both fracture angulation measurement methods, the standardized method had a much narrower 95% CI (0.954-0.987) compared with the raters' own measurement method (0.688-0.920), suggesting that the results for the standardized measurement method were more precise.

Good intrarater and moderate interrater reliability was seen for the NH classification. Subanalysis of each NH grade demonstrated good interrater agreement for grade I fractures (κ = 0.797; 95% CI, 0.713-0.881) but only moderate agreement for grades II (κ = 0.567; 95% CI, 0.483-0.651), III (κ = 0.444; 95% CI, 0.360-0.527), and IV fractures (κ = 0.498; 95% CI, 0.414-0.581).

The SH classification had the lowest agreement. Moderate intraobserver and fair interobserver reliability was achieved. Subanalysis of fracture types showed moderate interrater agreement for metaphyseal fractures (κ = 0.472; 95% CI, 0.388-0.556) but only fair agreement for SH I (κ = 0.227; 95% CI, 0.143-0.311) and II (κ = 0.265; 95% CI, 0.181-0.349) fractures.

DISCUSSION

There is no consensus in the literature with regard to indications for operative management of PHFs in adolescents.[2,6–12] Adolescents may benefit from surgical fixation of severely displaced fractures as they have less remaining growth and remodeling potential. Although some authors recommend operative treatment of NH grade III and IV fractures in patients 10 years and older with > 20 to 30 degrees of angulation,[2] other authors recommend surgery for patients 12 years and older with > 40 degrees of angulation and > 50% displacement.[7] Another study recommends surgery for patients 12 years and older with any amount of fracture angulation.[8] To establish consistent criteria for operative management of adolescent PHFs, it is important to first evaluate the reliability of measurements and classification systems that are commonly used to guide treatment recommendations.

The new standardized method of measuring angulation of adolescent PHFs that we describe in this study is similar to the head-shaft angle that is used to plan shoulder arthroplasties[29] and to assess fracture reduction after fixation of adult PHFs. The head-shaft angle is formed by a line perpendicular to the anatomic neck of the humerus and the axis of the humeral shaft.[30] As the proximal humeral physis is more clearly visible on radiographs than the anatomic neck in skeletally immature patients, we thought that adapting the measurement method for the epiphyseal-shaft angle for slipped capital femoral epiphysis would be more appropriate. In addition, Green et al[31] demonstrated excellent intraobserver and interobserver reliability of the epiphyseal-shaft angle in both normal hips and hips with slipped capital femoral epiphysis, whereas Gracitelli et al[20] found moderate intraobserver and fair interobserver agreement of the head-shaft angle in adult PHFs.

We achieved excellent intrarater and interrater reliability among all raters for the standardized method of measuring fracture angulation. In contrast, although the attending surgeons had excellent intraobserver agreement when raters used their own method of measuring fracture angulation, intraobserver agreement ranged from poor to good among the surgeons-in-training. Our standardized method is easy to learn and can be used reliably by both less experienced and experienced surgeons.

Multiple studies of adult PHFs have demonstrated low intrarater and interrater reliability of radiographic classification systems.[13–24] Similarly, we found good intrarater and moderate interrater reliability for the NH classification, which is the most commonly used system to

TABLE 1. Intrarater and Interrater Reliability Results for Fracture Angulation Measurement Methods and Classification Systems

	Intrarater Reliability	95% CI	Interrater Reliability	95% CI
Standardized angulation measurement method	0.959	0.864-0.991	0.973	0.954-0.987
Raters angulation measurement method	0.741	0.148-0.939	0.835	0.688-0.920
Neer and Horowitz classification	0.602	0.435-0.781	0.582	0.530-0.633
Salter-Harris classification	0.569	0.291-0.814	0.320	0.258-0.381

CI indicates confidence interval.

TABLE 2. Intraobserver Reliability Results for Each Rater

Rater*	Standardized Angulation Measurement Method	Raters Angulation Measurement Method	Neer and Horowitz Classification	Salter-Harris Classification
1	0.864	0.931	0.506	0.602
2	0.991	0.937	0.726	0.693
3	0.978	0.910	0.608	0.814
4	0.988	0.939	0.781	0.653
5	0.977	0.148	0.567	0.291
6	0.961	0.553	0.435	0.344
7	0.955	0.770	0.588	0.587

*Raters ranked in order of most to least experienced.

grade displacement of pediatric PHFs. Interobserver agreement decreased for the more displaced fractures. The most likely explanation for this is that NH grades II to IV are determined based on amount of displacement relative to humeral shaft width and NH grade I is determined based on an absolute measurement of displacement (ie, <5 mm).[3] The location on the humeral shaft at which the width was measured likely varied between raters. The NH classification system should be used with caution when determining the need for operative fixation of PHFs with more significant displacement.

Our study demonstrated moderate intrarater and fair interrater reliability for the SH classification. The presence or absence of metaphyseal extension of a proximal humerus physeal fracture can be difficult to detect on a plain radiograph. Multiple studies in adult PHFs have evaluated the use of computed tomography (CT) to improve fracture characterization,[13–16,19,24] although not all studies demonstrated improved intraobserver and interobserver agreement on CT compared with plain radiographs.[14,16,19] Three-dimensional CT has been shown to improve fracture classification in adult PHFs among less experienced raters compared with 2-dimensional CT.[13] The raters with less experience in our study had the lowest intrarater reliability for the NH and SH classifications. Although CT may improve fracture assessment, we do not recommend that patients with pediatric PHFs routinely undergo a CT scan, particularly as the SH classification is not used to guide treatment.

One limitation of our study is that only anteroposterior shoulder radiographs were evaluated. We may have achieved higher agreement with the SH classification if lateral views were included. However, lateral radiographs are not always obtained secondary to pain and difficulty with positioning when pediatric patients present with a PHF. In addition, the type of lateral view that is obtained is not consistent (eg, scapular Y vs. axillary lateral). Although we provided each rater with a description of the NH classification, we did not provide instructions on how to perform the grading. We may have found higher agreement for the NH classification if these instructions were provided. Other limitations are the small sample size, although the number of radiographs selected was the minimum number suggested by the power analysis. We did not include any SH III or IV PHFs, as these are rare fracture patterns. Lastly, we did not include radiographs of any patients under 10 years of age so we cannot necessarily extrapolate the results of this study to PHFs in younger patients.

In conclusion, our new standardized method for measuring angulation in adolescent PHFs achieved excellent intraobserver and interobserver agreement. We propose that our technique may be a more precise method of measuring angulation. This method is easy to learn and can be used reliably by surgeons with a wide range of experience. We suggest that this method be used in future studies that evaluate indications for operative management of adolescent PHFs.

REFERENCES

1. Landin LA. Epidemiology of children's fractures. *J Pediatr Orthop B*. 1997;6:79–83.
2. Fernandez FF, Eberhardt O, Langendorfer M, et al. Treatment of severely displaced proximal humeral fractures in children with retrograde elastic stable intramedullary nailing. *Injury*. 2008;39:1453–1459.
3. Neer CS II, Horwitz BS. Fractures of the proximal humeral epiphysial plate. *Clin Orthop Relat Res*. 1965;41:24–31.
4. Salter R, Harris W. Injuries involving the epiphyseal plate. *J Bone Joint Surg Am*. 1963;45:587–622.
5. Popkin CA, Levine WN, Ahmad CS. Evaluation and management of pediatric proximal humerus fractures. *J Am Acad Orthop Surg*. 2015;23:77–86.
6. Bahrs C, Zipplies S, Ochs BG, et al. Proximal humeral fractures in children and adolescents. *J Pediatr Orthop*. 2009;29:238–242.
7. Beaty JH. Fractures of the proximal humerus and shaft in children. *Instr Course Lect*. 1992;41:369–372.
8. Binder H, Schurz M, Aldrian S, et al. Physeal injuries of the proximal humerus: long-term results in seventy two patients. *Int Orthop*. 2011;35:1497–1502.
9. Burgos-Flores J, Gonzalez-Herranz P, Lopez-Mondejar JA, et al. Fractures of the proximal humeral epiphysis. *Int Orthop*. 1993;17:16–19.
10. Dameron TB Jr, Reibel DB. Fractures involving the proximal humeral epiphyseal plate. *J Bone Joint Surg Am*. 1969;51:289–297.
11. Dobbs MB, Luhmann SL, Gordon JE, et al. Severely displaced proximal humeral epiphyseal fractures. *J Pediatr Orthop*. 2003;23:208–215.
12. Hutchinson PH, Bae DS, Waters PM. Intramedullary nailing versus percutaneous pin fixation of pediatric proximal humerus fractures: a comparison of complications and early radiographic results. *J Pediatr Orthop*. 2011;31:617–622.
13. Berkes MB, Dines JS, Little MT, et al. The impact of three-dimensional CT imaging on intraobserver and interobserver reliability of proximal humeral fracture classifications and treatment recommendations. *J Bone Joint Surg Am*. 2014;96:1281–1286.
14. Bernstein J, Adler LM, Blank JE, et al. Evaluation of the Neer system of classification of proximal humeral fractures with computerized tomographic scans and plain radiographs. *J Bone Joint Surg Am*. 1996;78:1371–1375.
15. Bruinsma WE, Guitton TG, Warner JJ, et al. Science of Variation Group. Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am*. 2013;95:1600–1604.
16. Foroohar A, Tosti R, Richmond JM, et al. Classification and treatment of proximal humerus fractures: inter-observer reliability and agreement across imaging modalities and experience. *J Orthop Surg Res*. 2011;6:38.
17. Majed A, Macleod I, Bull AM, et al. Proximal humeral fracture classification systems revisited. *J Shoulder Elbow Surg*. 2011;20:1125–1132.
18. Sidor ML, Zuckerman JD, Lyon T, et al. The Neer classification system for proximal humeral fractures. An assessment of interobserver

reliability and intraobserver reproducibility. *J Bone Joint Surg Am*. 1993;75:1745–1750.
19. Sjoden GO, Movin T, Guntner P, et al. Poor reproducibility of classification of proximal humeral fractures. Additional CT of minor value. *Acta Orthop Scand*. 1997;68:239–242.
20. Gracitelli ME, Dotta TA, Assuncao JH, et al. Intraobserver and interobserver agreement in the classification and treatment of proximal humeral fractures. *J Shoulder Elbow Surg*. 2017;26:1097–1102.
21. Kristiansen B, Andersen UL, Olsen CA, et al. The Neer classification of fractures of the proximal humerus. An assessment of interobserver variation. *Skeletal Radiol*. 1988;17:420–422.
22. Papakonstantinou MK, Hart MJ, Farrugia R, et al. Interobserver agreement of Neer and AO classifications for proximal humeral fractures. *ANZ J Surg*. 2016;86:280–284.
23. Handoll HH, Brealey SD, Jefferson L, et al. Defining the fracture population in a pragmatic multicentre randomised controlled trial: PROFHER and the Neer classification of proximal humeral fractures. *Bone Joint Res*. 2016;5:481–489.
24. Brunner A, Honigmann P, Treumann T, et al. The impact of stereo-visualisation of three-dimensional CT datasets on the inter- and intraobserver reliability of the AO/OTA and Neer classifications in the assessment of fractures of the proximal humerus. *J Bone Joint Surg Br*. 2009;91:766–771.
25. Southwick WO. Osteotomy through the lesser trochanter for slipped capital femoral epiphysis. *J Bone Joint Surg Am*. 1967;49:807–835.
26. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
27. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76:378–382.
28. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
29. Pearl ML. Proximal humeral anatomy in shoulder arthroplasty: implications for prosthetic design and surgical technique. *J Shoulder Elbow Surg*. 2005;14(suppl S):99S–104S.
30. Malavolta EA, Assuncao JH, Pagotto RA, et al. The rotation of the humeral head does not alter radiographic evaluation of the head-shaft angle. *J Shoulder Elbow Surg*. 2016;25:543–547.
31. Green DW, Mogekwu N, Scher DM, et al. A modification of Klein's Line to improve sensitivity of the anterior-posterior radiograph in slipped capital femoral epiphysis. *J Pediatr Orthop*. 2009;29:449–453.