Reliability of radiographic measurement of fracture angulation in children with femoral shaft fractures

Janice L. Owen, BScPT;*† Derek Stephens, MSc;* James G. Wright, MD*†

Introduction: Clinical decisions are based on x-ray measures of radiographs. However, the reliability of assessing the angular measurement of fractured femurs in children is unknown. The purpose of this study was to determine the intra- and interrater reliability of an x-ray assessment of femoral shaft fracture angulation in children. **Methods:** On 2 occasions, 3 raters evaluated 30 radiographs of children aged 4 to 10 years, 3 to 24 months after a femoral fracture. The radiographs were evaluated by an orthopedic surgeon, an orthopedic fellow and a research assistant, using a standardized methodology. Reliability was assessed with the intraclass correlation coefficient (ICC), an index of concordance that ranges from 0–1, where > 0.75 indicates excellent concordance. **Results:** The ICC for the intrarater ranged from 0.75 to 0.97. The interrater reliability ICC for anterior or posterior angulation was 0.91 (95% confidence interval [CI] 0.84–0.95) and for varus or valgus angulation was 0.90 (95% CI 0.79–0.95). A change in measurement of more than 5 degrees exceeds the variation in measurement and, therefore, can be attributed to a true shift in fracture position. **Conclusions:** With specific standardized protocols, radiographic assessment of femoral fractures is reliable, irrespective of the examiner's level of experience.

Introduction : Les décisions cliniques se fondent sur des évaluations radiologiques de radiographies. On ne connaît toutefois pas la fiabilité de l'évaluation de la mesure angulaire des fractures du fémur chez les enfants. Cette étude visait à déterminer le coefficient d'objectivité personnelle et entre évaluateurs d'une évaluation radiologique de l'angle d'une fracture du corps du fémur chez les enfants. **Méthodes :** À deux occasions, trois évaluateurs ont évalué 30 radiographies d'enfants âgés de 4 à 10 ans, de 3 à 24 mois après une fracture du fémur. Les radiographies ont été évaluées par un chirurgien orthopédique, par un associé en orthopédie et par un adjoint de recherche, à l'aide d'une méthodologie normalisée. On a évalué la fiabilité au moyen du coefficient de corrélation intraclasse (intraclass correlation coefficient, CCI), indice de concordance qui varie de 0 à 1, où > 0,75 indique une excellente concordance. **Résultats :** Dans le cas de l'objectivité de l'évaluateur, le CCI a varié de 0,75 à 0,97. Le CCI de l'objectivité entre évaluateurs dans le cas d'une déviation antérieure ou postérieure s'est établi à 0,91 (intervalle de confiance [IC] à 95 %; 0,84–0,95) et à 0,90 (IC à 95 %; 0,79–0,95) dans celui d'une déviation en varus ou en valgus. Un changement de mesure de plus de cinq degrés dépasse la variation de la mesure et peut donc être attribué à un déplacement réel de la position de la fracture. **Conclusions :** En présence de protocoles normalisés spécifiques, l'évaluation radiographique des fractures du fémur est fiable, quel que soit le niveau d'expérience de l'évaluateur.

Femoral fractures are common[1,2] and expensive.[3–5] Numerous options are available for treatment, including traction, casting, external fixation and flexible intramedullary nails. Treatment is influenced by many factors and varies widely within the Pediatric Orthopedic Society of North America.[6] Irrespective of treatment, clinical decisions are made using a variety of information, including physical assessment and interpretation of radiographs. Initially, radiographic measurements are used to assess satisfactory fracture position.[7] One of the major components of satisfactory position is the angulation of the fracture fragments. Subsequently, a change between radiographic measurements may indicate a shift in fracture position that might require such treatment as cast wedging, adjustment of a fixator or change to a different modality. However, a change between x-ray measurements may also be due to

Correspondence to: Dr. James G. Wright, The Hospital for Sick Children, 555 University Ave., Toronto ON M5G 1X8; fax 416 813-6433; james.wright@sickkids.ca

variability of measurement, in which case, no change in treatment would be indicated. Thus, measures need to be as reliable as possible to attribute change in measures to true clinical change.[8] The purpose of this study was to evaluate the reliability of angulation measurements from radiographs of children with femoral fractures.

Methods

The radiographs used in this study were those of children aged 4 to 10 years, who were admitted to 4 pediatric hospitals, had consented to participate in a prospective trial and had received either early application of hip spica or external fixation.[9] The study received research ethics board approval in each site, and we obtained informed consent from the participants' parents or guardians. Radiographs consisting of anterior or posterior and lateral views were obtained at injury and at 3, 9, 15 and 24 months postinjury.

Three examiners, including a staff pediatric orthopedic surgeon, a pediatric orthopedic fellow and a research assistant, reviewed the radiographs. The research assistant had no prior experience evaluating radiographic fracture positions.

The measurement techniques, discussed below, were decided a priori. Although all examiners knew their measurements were part of a reliability study, no joint session or common discussions occurred. Each rater interpreted the radiographs independently and was allowed unrestricted time on each set of radiographs. Raters were blinded to their own and to other raters' evaluations. Radiographs were numbered sequentially with all data regarding patient identity or study site removed or blocked out.

All x-ray measurements were performed with a soft-lead pencil and the same protractor. After the measurements had been recorded, the markings were erased with isopropyl alcohol. Each rater evaluated the angular position of the fractures on the radiographs on 2 separate occasions,

with at least a 14-day interval between measurements. We chose 14 days to minimize the recall of prior measurements. During the subsequent evaluations, radiographs were presented in random order to guard against any recall bias from the first interpretation.

For the measurement of fracture positions, radiographs were oriented to the upright position, and 2 lines were drawn so they intersected in the centre of the intramedullary canal, parallel to the shaft of the proximal segment of the femoral shaft, and in the centre of the intramedullary canal, parallel to the shaft of the distal segment of the femoral shaft. This technique is essential to avoid using the centre of the rounded-off apex of the fracture, thereby falsely minimizing the angular deformity.[10] The protractor was aligned with the zero mark against the proximal segment. The angle reported was then measured as varus or valgus angulation, using the anterior or posterior view, and the anterior or posterior angulation was measured using the lateral view.

Reliability was quantified with SPSS 10.1 for Windows for both intrarater and interrater reliability, using the intraclass correlation coefficient (ICC). The ICC, a test of concordance or agreement for continuous data, such as measurement angles, ranges from 0 to 1, with ICC greater than 0.75 indicating excellent reliability.[11] The random effects model allows for extrapolation to all raters and is useful for the intrarater and interrater trials.[8] To assess interrater reliability, the average ICC over 2 occasions was used.

The sample size for the study was estimated for a random effects ICC. Assuming an alpha level of 0.05, beta level of 0.20, ICC of 0.8 and taking 2 measurements per subject, the estimated sample size was 30 patients.[12]

To quantify the variability of measurements obtained by the 3 raters, we performed limits of agreement. In this technique, described by Bland and Altman,[13] the limits of agree-

ment are the mean of the difference plus or minus 2 standard deviations (SDs). That is, 95% limits of agreement = $2s$. Change in fracture position beyond the limits of agreement can be more confidently attributed to true change rather than to simple variability of measurement.

We performed correlations to examine the covariance between pairs of raters for both varus or valgus and anterior or posterior angles.[14]

Results

Of the 30 patients included in this study, 17 were male and 13 were female. Sixteen received hip spica and 14 received external fixation (all casts and fixators had been removed at the time of the radiographs). Of the 30 patients, 5 had spiral, 11 had oblique and 5 had spinal fractures. Patients were obtained from all 4 centres and thus are typical of children with femoral fractures in the age range of 4 to 10 years. The definition of malunion, the primary outcome of the study, was defined as more than 10° of varus or valgus, or 15° of anterior or posterior angulation. The results of the intrarater ICC are presented in Table 1. The research assistant achieved the best rating, with an ICC of 0.97 for both anterior or posterior and varus or valgus angulation. The staff surgeon's initial ICC was 0.86 and 0.55 for anterior or posterior and varus or valgus angulation, respectively. After another review was performed, the ICCs were 0.90 and 0.75, respectively, using the first and third x-ray readings.

Interrater reliability for anterior or posterior angulation was 0.91 (95% confidence interval [CI] 0.84–0.95) and varus or valgus angulation was 0.90 (95% CI, 0.79–0.95). The overall reliability for both varus or valgus and anterior or posterior angulation was excellent.

The limits of agreement, as shown in Table 2, demonstrate that the 3 raters, regardless of the angle, were almost identical. Agreement between

the research assistant and the staff orthopedic surgeon was the strongest.

The Pearson correlation coefficients, as displayed in Table 3, were positive and extremely strong for all pairs of rater comparisons with $r = 0.9$ ($p = 0.01$).

Discussion

Measurement of fracture angulation is fundamental to fracture care. For all fractures, including femoral fractures, radiographs are used to determine an appropriate reduction and to subsequently provide criteria for malunion. Reliability is a basic prerequisite of all clinic measurements and is defined as obtaining the same measurement by the same or by a different clinician.[8] Variability in measurement can arise from 3 sources: the patient, the procedure and the clinician.[8]

This paper focuses on the intrarater and interrater reliability of clinicians' reading of radiographs. Physicians who see patients with a femoral fracture should know the precision and reproducibility of the angles that are measured. There is no universally accepted definition of an acceptable reduction for femoral fractures in children. However, in the randomized trial, we used more than 10° of varus or valgus or 15° of anterior or posterior angulation.[9] To guide clinicians on what constitutes a true change in fracture position, we used the limits of agreement. The results of this study suggest that a change in angulation of more than 5° can be attributed to a true shift in fracture position. This may seem intuitive, but it should be reassuring to clinicians who frequently assess patients. When fractures move beyond this limit, and if the progression is into the unacceptable range, then treatment, such as cast wedging or fixation adjustment, may be required.

The least experienced rater in this study performed as well or better than the more experienced raters. This indicates that skill in rating radiographs is not dependent on clinical experience. Although the initial reliability of the staff was moderate, the absolute variance was relatively small. Remeasurement demonstrates that precision can be improved simply through awareness.

The most important source of variability not assessed in this study was the procedure involved in taking x-rays, specifically patient positioning. For the appraisal of angular deformity, a potential source of variability is the effect of rotation of the limb. Given the radiation exposure, repeated x-rays of patients at the same time is not ethical. However, previous research in cadaveric models has shown that rotation of up to 10° has a minimal effect on angular deformity.[15] Although the results of this study show high levels of reliability for the radiographic assessment of fracture position, this estimate provides only a minimal level of reliability, given the inability to perform repeat x-rays on the same visit. The clinicians involved had neither special training nor a joint session to practice their technique. Standardizing the

Table 1

Intrarater reliability

Rater	Anterior or posterior angulation		Varus or valgus angulation	
	ICC	95% CI	ICC	95% CI
Orthopedic fellow	0.91	10.081–0.95	0.88	0.77–0.94
Research assistant	0.97	0.94–0.99	0.97	0.93–0.99
Staff surgeon, initial	0.86	0.73–0.93	0.55	0.25–0.76
Staff surgeon, repeat	0.90	0.80–0.96	0.75	0.54–0.87

ICC = intraclass correlation coefficient; CI = confidence interval.

Table 2

Pairwise comparisons of raters

Deformity: raters	Mean difference (SD), 95% limits of agreement
Varus/valgus:	
R1–R2	0.03 (1.7), −3.3–3.4
R1–R3	0.2 (1.8), −3.3–3.7
R2–R3	0.1 (1.8), −3.5–3.8
Anterior/posterior:	
R1–R2	-0.3 (2.6), −5.6–5.0
R1–R3	0.1 (2.3), −4.5–4.6
R2–R3	0.4 (2.3), −4.0–5.0

SD = standard deviation; R = rater.
R1 = orthopedic fellow; R2 = research assistant; R3 = staff orthopedic surgeon.

Table 3

Correlation matrices for pairs of raters

Deformity: rater	Pearson correlation coefficient (and *p* value)		
	Orthopedic fellow	Research assistant	Staff surgeon
Varus/valgus:			
Orthopedic fellow	1 (0)	NA	NA
Research assistant	0.916 (0.01)	1 (0)	NA
Staff surgeon	0.915 (0.01)	0.890 (0.01)	1 (0)
Anterior/posterior:			
Orthopedic fellow	1 (0)	NA	NA
Research assistant	0.901 (0.01)	1 (0)	NA
Staff surgeon	0.919 (0.01)	0.925 (0.01)	1 (0)

technique and tools controlled the measurement process. If the results of this study are to be generalized to other settings, clinicians should use a similar method of determining where to place the lines, as described in this study.

In conclusion, the measurement of femoral fracture angulation from radiographs, using a standardized technique, is reliable regardless of the clinician's experience. With the method described in this study, clinicians can be confident that a true change in fracture angulation of 5° probably constitutes a true change in fracture position.

References

1. Nafei A, Teichert G, Mikkelsen SS, et al. Femoral shaft fractures in children: an epidemiological study in a Danish urban population, 1977-86. *J Pediatr Orthop* 1992;12:499-502.
2. Hinton RY, Lincoln A, Crockett MM, et al. Fractures of the femoral shaft in children. *J Bone Joint Surg Am* 1999(81):500-9.
3. Newton PO, Mubarak SJ. Financial aspects of femoral shaft fracture treatment in children and adolescents. *J Pediatr Orthop* 1994;14:508-12.
4. Coyte PC, Bronskill SE, Hirji ZZ, et al. Economic evaluation of 2 treatments for pediatric femoral shaft fractures. *Clin Orthop Relat Res* 1997;(336):205-15.
5. Buechsenschuetz KE, Mehlman CT, Shaw KJ, et al. Femoral shaft fractures in children: traction and casting versus elastic stable intramedullary nailing. *J Trauma* 2002;53:914-21.
6. Sanders JO, Browne RH, Mooney JF, et al. Treatment of femoral fractures in children by pediatric orthopedists: results of a 1998 survey. *J Pediatr Orthop* 2001;21:436-41.
7. Staheli LT. Fractures of the shaft of the femur. In: Rockwood CA, Wilkins KE, King RE, eds. *Fractures in Children*. 3rd ed. New York: Lippincott; 1991. p. 1121-63.
8. Wright JG, Feinstein AR. Improving the reliability of orthopaedic measurements. *J Bone Joint Surg Br* 1992;74:287-91.
9. Wright JG, Wang EEL, Owen JL, et al. Treatments for paediatric femoral fractures: a randomized trial. *Lancet* 2005;365:1153-8.
10. Wallace ME, Hoffman EB. Remodelling of angular deformity after femoral shaft fractures in children. *J Bone Joint Surg Br* 1992;74:765-9.
11. Fleiss J. Reliability of measurement. In: *The design and analysis of clinical experiments*. New York: John Wiley and Sons; 1986. p. 1-31.
12. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med* 1987;6:441-8.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.
14. Portney LG, Watkins MP. Correlation. In: *Foundations of clinical research: applications to practice*. 2nd ed. Upper Saddle River: Prentice Hall Health; 2000. p. 491-508.
15. Wright JG, Treble N, Feinstein AR. Measurement of lower limb alignment using long radiographs. *J Bone Joint Surg Br* 1991;73:721-3.