

Interobserver and intraobserver reliability of radiographic evidence of bone healing at osteotomy sites

Ashish Anand, David S. Feldman, Raviraj J. Patel, Wallace B. Lehman, Harold J.P. van Bosse, Mohammad I. Badra and Debra A. Sala

Plain radiographs are commonly used to evaluate the degree of bone healing after an osteotomy and the application of an external fixator. The purpose of the study was to assess intraobserver and interobserver reliability in determining bone healing, defined as bridging callus across three of four cortices, of osteotomy sites on radiographs. Substantial intraobserver reliability and a high intraobserver percentage agreement were found. Interobserver reliability was moderate and interobserver percentage agreement was less than half for agreement between all involved orthopaedic surgeons. The lower reliability across surgeons suggests that the determination of the extent of the bone healing is subjective. *J Pediatr Orthop B* 15:271–272 © 2006 Lippincott Williams & Wilkins.

Introduction

Plain radiographs are the present standard for assessing the degree of bone healing following an osteotomy and application of an external fixator in deformity correction or limb lengthening. The ability to accurately determine the point when an osteotomy is adequately healed is important to prevent fracture, deformity or nonunion at the osteotomy site. Additionally, the timing of fixator removal is based on the orthopedic surgeon's judgment of the solidity of the bone regenerate or callus.

Orbay *et al.* [1] demonstrated that more extensive radiographic healing was associated with greater mechanical strength of the bone regenerate in canine tibial lengthening with an Ilizarov external fixator. In a clinical study, correlation to diminished refracture rate was seen in pediatric femoral fractures in which the fracture had bridging callus in three cortices [2]. These authors suggested that the fixator should be removed only after the bridging callus is observed radiographically across a minimum of three cortices.

Physicians' ability to reliably assess the number of continuous cortices across an osteotomy site and whether healing was sufficient for fixator removal using antero-posterior and lateral radiographs was evaluated in patients undergoing lengthening with external fixators [3]. Intraobserver reliability was low and interobserver reliability was only slight for the number of healed cortices identified. Intraobserver reliability was moderate and interobserver reliability was only fair regarding the decision about readiness for fixator removal.

Journal of Pediatric Orthopaedics B 2006, 15:271–272

Keywords: bone healing, osteotomy, radiographs, reliability

Pediatric Orthopaedic Surgery, Center for Children, NYU Hospital for Joint Diseases, New York, New York, USA

Correspondence and requests for reprints to David S. Feldman, MD, Chief, Pediatric Orthopaedic Surgery, Center for Children, NYU Hospital for Joint Diseases, 301 East 17th Street, New York, New York 10003, USA.
Tel: +1 212 533 5310; fax: +1 212 533 5345;
e-mail: david.feldman@med.nyu.edu

Study conducted at the Center for Children, NYU Hospital for Joint Diseases, New York, New York, USA.

The purpose of this study was to assess the intraobserver and interobserver reliability in determining bone healing, defined as bridging callus across three of four cortices, of osteotomy sites on plain radiographs.

Methods

Forty-seven radiographs taken at a mean of 78.85 days (range 12–232 days) post-fixator application were randomly selected from a sample of 12 patients, mean age 12.9 years (range 5–35 years), who had undergone a tibial osteotomy and application of an external fixator. Twenty (42.6%) of the radiographs were from five patients who had corrective osteotomies and 27 (57.4%) of the radiographs were from seven patients who had corrective and lengthening osteotomies.

Bone healing was defined as bridging callus visible across at least three of four cortices of the osteotomy site. Before the initial measurement of the radiographs, each surgeon was provided with a written copy of the following definition of what constituted bone healing: cortical bridging, defined as the disappearance of the cortical interruption of the osteotomy site as a result of callus formation, visible across at least three cortices.

Physicians were blinded to patient's names and dates of the radiographs. Four attending pediatric orthopedic surgeons and one pediatric orthopedic surgery fellow reviewed the radiographs and scored them as healed or not healed. At a mean of 11.2 days (range 7–23 days) following the first scoring session, the same group of

Table 1 Statistical analysis of intraobserver reliability

	Surgeon 1	Surgeon 2	Surgeon 3	Surgeon 4	Surgeon 5	Mean
Kappa coefficient	0.63	0.56	0.46	0.83	0.79	0.65
Percentage agreement (%)	93.6	87.2	85.1	91.5	89.4	89.4

Table 2 Interobserver κ coefficients

	Surgeon 1	Surgeon 2	Surgeon 3	Surgeon 4	Surgeon 5
Surgeon 1		0.41	0.46	0.24	0.18
Surgeon 2			0.69	0.52	0.40
Surgeon 3				0.47	0.29
Surgeon 4					0.66
Mean	0.32	0.51	0.48	0.47	0.38

orthopedic surgeons reviewed the same radiographs and scored them a second time as healed or not healed.

Statistical analysis

Kappa coefficient is a measure of the amount of agreement in nominal data between two observers and includes a correction for chance agreement. The coefficient can range from +1.00 to -1.00, with +1.00 indicating perfect agreement, 0 only chance agreement and any negative values indicating less than chance agreement [4]. The intraobserver reliability was calculated by comparing the responses made in sessions 1 and 2 for each surgeon by using the κ coefficient and percentage agreement. By comparing the responses made in session 1 by each surgeon with those of the other surgeons, the interobserver reliability was determined using the κ coefficient. Additionally, interobserver percentage agreement was calculated for when four of the five surgeons agreed and when all five surgeons agreed. The interpretation of the κ coefficients was based on the criteria of Landis and Koch [5]. All analyses were carried out using SPSS 10.0 (SPSS Inc., Chicago, Illinois, USA).

Results

For intraobserver reliability, the mean κ coefficient was 0.65 (range 0.46–0.83) (Table 1), demonstrating substantial intraobserver reliability. The mean intraobserver percentage agreement between sessions 1 and 2 was 89.4% (range 85.1–93.6%) (Table 1).

For interobserver reliability, the mean κ coefficient was 0.43 (range 0.18–0.69) (Table 2), indicating moderate interobserver reliability. The interobserver percentage agreement of four out of five surgeons in session 1 was 72.3% and five out of five surgeons in session 1 was 48.9%.

For the radiographs that were taken < 60 days after fixator application, 16 of 19 (84.2%) were scored unanimously as not healed. For radiographs taken > 60 days after fixator application, 7 of 28 (25%) were scored unanimously as not healed or healed.

Discussion

During limb lengthening or deformity correction using external fixation, radiographs are typically used by surgeons after an osteotomy to monitor the progress of osteogenesis and to evaluate the adequacy of bone healing for fixator removal. Applying the commonly suggested criterion for bone healing, that is, bridging callus across three of four cortices, five orthopedists in the present study agreed in less than half of the cases. When this was reduced to agreement by four of the five surgeons, the agreement rate improved to almost 75%. Starr *et al.* [3] had suggested to the examiners in their study to use the same criterion as the determinant of the time for removal of the fixator. In one-quarter of the cases in which they had indicated healing across three or four cortices, however, the examiners did not feel that healing was sufficient for fixator removal. These authors reported that additional factors, such as the location of the osteotomy and the quality of the callus, are taken into consideration when deciding on whether to remove a fixator.

In comparing the κ coefficients of the present study and those for time for fixator removal found by Starr *et al.* [3], our higher values for both intraobserver (0.65 vs. 0.46, respectively) and interobserver (0.43 vs. 0.35, respectively) could be attributed to the post-operative date of the radiographs. Our radiographs spanned the entire period from shortly after fixator application to removal while, in the study of Starr *et al.* [3], they were limited to the period of consolidation. In the time interval from the date of osteotomy to completion of distraction, which was only included in our study, agreement that the site is not healed is more likely and, therefore, would increase the magnitude of the coefficients.

The lack of reliability between surgeons suggests that a high level of subjectivity is involved in determining the extent of the bone regenerate on radiographs.

References

- Orbay JL, Frankel VH, Finkle JE, Kummer FJ. Canine leg lengthening by the Ilizarov technique. *Clin Orthop* 1992; **278**:265–273.
- Skaggs DL, Leet AI, Money MD, Shaw BA, Hale JM, Tolo VT. Secondary fractures associated with external fixation in pediatric femur fractures. *J Pediatr Orthop* 1999; **19**:582–586.
- Starr KA, Fillman R, Raney EM. Reliability of radiographic assessment of distraction osteogenesis site. *J Pediatr Orthop* 2004; **24**:26–29.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; **20**:37–46.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**:159–174.