Interobserver and Intraobserver Reliability in Lower-Limb Deformity Correction Measurements

David S. Feldman, MD, Eric R. Henderson, MD,§ Harlan B. Levine, MD,* Philip L. Schrank, MD,* Kenneth J. Koval, MD,† Raviraj J. Patel, BS,‡ Daniel B. Spencer, BS,* Debra A. Sala, MS, PT,* and Kenneth A. Egol, MD**

Abstract: Planning for surgical correction of lower-limb deformity requires assessment of the character and extent of the deformity. Deformity measurements are defined; however, the reliability of these measurements has not been evaluated. This study was conducted to assess the interobserver and intraobserver reliability of lower extremity deformity measurements in the frontal and sagittal planes. Anteroposterior and lateral lower extremity radiographs were evaluated using Paley technique. Statistical analysis included intraclass correlation coefficient (2,1), median absolute difference, range, and agreement within 3 and 5 degrees. Reliability was good to very good for all measurements except for the anterior distal tibial angle, which had moderate reliability. Intraobserver reliability was higher than interobserver reliability, and measurements in the frontal plane had better reliability than measurements in the sagittal plane. Overall, these measurements are a reliable method of assessing lower extremity deformity and should be used to guide treatment and monitor outcome.

Key Words: deformity correction, lower extremity, lower limb, measurement, reliability, femur, tibia

(*J Pediatr Orthop* 2007;27:204–208)

Lower-limb deformity is a disorder which, depending on severity, may be remedied with surgical or nonsurgical intervention. Paley et al[1] and Paley and Tetsworth[2,3] described deformity correction planning for uniapical and multiapical deformities. Treatment planning requires assessment of the character and extent of the deformity in anteroposterior (AP) and lateral radiographs.

Deformity measurement methods are defined; however, the process has inherent subjectivity. These measurements are indispensable in guiding patient care, particularly with some newer devices that require measurement input for correction. Therefore, it is important that there should be a significant degree of consistency and reproducibility of the results. The present study was intended to assess the interobserver and intraobserver reliability of lower extremity deformity assessment in the frontal and sagittal planes.

METHODS

Deformity measurements were performed on standard AP and lateral radiographs using the technique described by Paley et al,[1] and Paley and Tetsworth,[2,3] (Fig. 1)[4]. Two sets of radiographs, 65 AP films and 76 lateral films, were compiled.

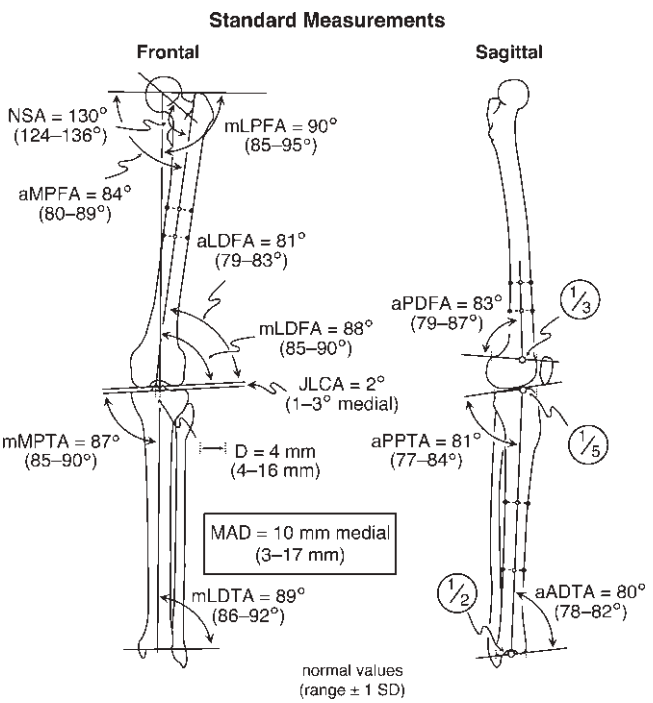

FIGURE 1. Standard lower extremity deformity measurements. Frontal plane abbreviations are the following: NSA = neck shaft angle; mLPFA = mechanical lateral proximal femoral angle; aMPFA = anatomic medial proximal femoral angle; aLDFA = anatomic lateral distal femoral angle; mLDFA = mechanical lateral distal femoral angle; JLCA = joint line congruency angle; mMPTA = mechanical medial proximal tibial angle; D = distance; MAD = mechanical axis deviation; mLDTA = mechanical lateral distal tibial angle. Sagittal plane abbreviations are the following: aPDFA = anatomic posterior distal femoral angle; aPPTA = anatomic posterior proximal tibial angle; aADTA = anatomic anterior distal tibial angle. Reprinted with permission from Springer-Verlag Berlin Heidelberg 2003.

From the *New York University Hospital for Joint Diseases, New York, NY; †Dartmouth Hitchcock Medical Center, Lebanon, NH; ‡New York University, New York, NY; and §Department of Orthopaedic Surgery, University of South Florida, Tampa, FL.

None of the authors received financial support for this study.

Reprints: David S. Feldman, MD, Pediatric Orthopaedic Surgery, Center for Children, New York University Hospital for Joint Diseases, 301 East 17th St, New York, NY 10003. E-mail: david.feldman@med.nyu.edu.

Frontal plane measurements performed were anatomic medial proximal femoral angle (aMPFA), anatomic lateral distal femoral angle (aLDFA), mechanical medial proximal tibial angle (mMPTA), and mechanical lateral distal tibial angle (mLDTA) (Fig. 2). Sagittal plane measurements were anatomic posterior distal femoral angle (aPDFA), anatomic posterior proximal tibial angle (aPPTA), and anatomic anterior distal tibial angle (aADTA) (Fig. 2).

Three physicians individually measured each film on 2 occasions. At each viewing, the physicians were given the opportunity to review the standardized measurement method of Paley et al.[1] A pediatric orthopaedist, an orthopaedic traumatologist, and a chief resident in orthopaedics measured the AP films. The same pediatric orthopaedist, a different orthopaedic traumatologist, and a different chief resident in orthopaedic surgery measured the lateral films. The physicians were required to perform their measurements isolated from one another. A minimum of 3 weeks between measurements was required. The same goniometer, straight-edged ruler, and protractor were supplied to each investigator. Markings on the radiographs were made with a soft-lead pencil with a 0.5 mm diameter lead tip. Angles were measured in degrees. All traces of previous measurements were erased before each viewing with an alcohol swab. Radiographs were assessed before each viewing to ensure that the integrity of image was not compromised after erasure of the previous measurements.

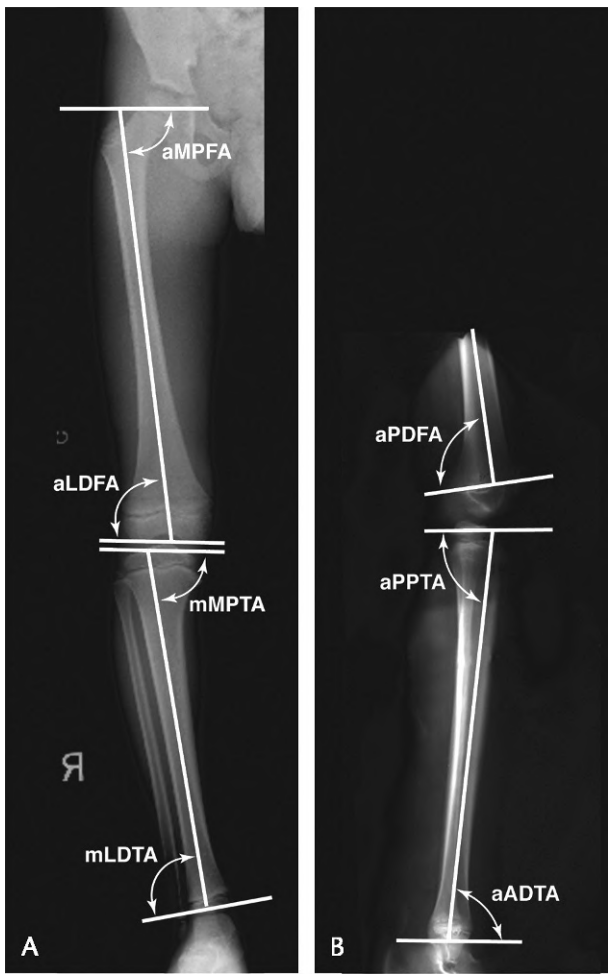

FIGURE 2. A, Frontal plane measurements investigated in the current study are the following: aMPFA, aLDFA, mMPTA, and mLDTA. B, Sagittal plane measurements investigated in the current study are the following: aPDFA, aPPTA, and aADTA.

TABLE 1. ICCs for Frontal Plane Measurements

	Intraobserver				Interobserver
Measurement	Observer 1	Observer 2	Observer 3	Mean	
aMPFA (n = 60)	0.95	0.89	0.96	0.93	0.91
aLDFA (n = 60)	0.95	0.90	0.96	0.94	0.89
mMPTA (n = 60)	0.83	0.65	0.86	0.78	0.77
mLDTA (n = 60)	0.79	0.77	0.84	0.80	0.70
Mean	0.88	0.80	0.91	0.86	0.82

Interobserver and intraobserver reliability were quantified using the intraclass correlation coefficient (ICC) (2,1). Reliability was scored based on the criteria by Altman[5] (very good, 0.81–1.00; good, 0.61–0.80; moderate, 0.41–0.60; fair, 0.21–0.40; and poor, ≤0.20).

For the interobserver data, the absolute difference between the measurements of the 3 pairs of observers (1 and 2, 1 and 3, and 2 and 3) was calculated for each angle. For the intraobserver data, the absolute difference between the measurements of sessions 1 and 2 of the individual observers was determined for each angle. Absolute difference, median, SD, and range values were calculated for both interobserver and intraobserver data sets.

Interobserver agreement of less than or equal to 3 and 5 degrees was said to exist if the absolute difference values for all 3 observer pairs was less than or equal to 3 and 5 degrees for a given radiograph. The percentage of interobserver agreement was determined for each of the angles measured. Intraobserver agreement of less than or equal to 3 and 5 degrees occurred if the absolute difference value between sessions 1 and 2 was less than or equal to 3 and 5 degrees for an individual observer. The

TABLE 2. ICCs for Sagittal Plane Measurements

	Intraobserver				Interobserver
Measurement	Observer 1	Observer 2	Observer 3	Mean	
aPDFA (n = 46)	0.90	0.79	0.76	0.82	0.71
aPPTA (n = 64)	0.94	0.92	0.86	0.91	0.86
aADTA (n = 62)	0.80	0.60	0.40	0.60	0.52
Mean	0.88	0.77	0.67	0.78	0.70

TABLE 3. Median Absolute Difference and Range for Frontal Plane Measurements

	Intraobserver (degrees)				Interobserver (degrees)
Measurement	**Observer 1**	**Observer 2**	**Observer 3**	**Mean**	
aMPFA (n = 60)	2.0 (0–14)	2.5 (0–23)	2.0 (0–10)	2.2	3.0 (0–21)
aLDFA (n = 60)	1.0 (0–7)	2.0 (0–13)	1.0 (0–6)	1.3	2.0 (0–19)
mMPTA (n = 60)	1.0 (0–30)	3.0 (0–28)	1.0 (0–29)	1.7	2.0 (0–29)
mLDTA (n = 60)	2.0 (0–26)	2.0 (0–18)	2.0 (0–19)	2.0	2.0 (0–22)
Mean	1.5	2.4	1.5	1.8	2.3

percentage of intraobserver agreement was calculated for each of the angles measured for each observer.

Each observer was given the ability to exclude a film from the study if it was deemed to be of poor quality. Five AP films were deemed unacceptable by at least 1 observer and were removed from the study. Ten lateral films were deemed unacceptable by at least 1 observer and were removed from the study. From the remaining lateral films, the distal femur was deemed unsuitable for measurement by at least 1 observer in 20 additional films, leaving 46 distal femoral measurements for consideration. Likewise, 2 proximal tibias and 4 distal tibias were considered unsuitable for measurement by at least 1 observer and were therefore excluded from the study, leaving 64 proximal tibial and 62 distal tibial measurements for statistical analysis.

All data analyses were performed using SPSS 10.0 (SPSS, Inc, Chicago, Ill).

Institutional review board approval was obtained for this project.

RESULTS

Interobserver frontal plane measurement reliability was highest at the hip and lowest at the ankle, with ICC (2,1) values scored as very good at the proximal and distal femur by Altman criteria[5] and good at the proximal and distal tibia (Table 1). Intraobserver reliability was also higher for the femur and lower around the tibia, with very good scores for all measurements proximal to the knee joint. Frontal plane intraobserver reliability scores were generally higher than the interobserver scores (Table 1).

Interobserver ICCs for sagittal plane measurements varied more than those for the frontal plane measurements. These values showed very good (aPPTA), good (aPDFA), and moderate (aADTA) reliability (Table 2). Again, the lowest interobserver ICC was seen in the measurements of the distal tibia. As with the frontal plane values, intraobserver ICCs for sagittal plane measurements were higher than interobserver ICCs, with very good intraobserver ICCs for aPDFA and aPPTA and a moderate intraobserver ICC for the aADTA (Table 2).

No statistically significant difference in reliability was found between age groups for either frontal or sagittal plane measurements.

The interobserver and intraobserver median absolute difference values indicate that most measurements for a given radiograph were within 3 degrees of one another (Tables 3 and 4). The intraobserver median absolute difference values were all less than the interobserver median absolute difference values, indicating higher precision within repeated measurements by a single observer than within measurements by multiple observers; this finding corroborates the higher intraobserver ICC (2,1) scores compared with interobserver ICC (2,1) scores.

Better intraobserver precision was also reflected in the lower intraobserver range values compared with the interobserver range values. The large range of measured values, up to 30 degrees, for a given radiograph indicate that a high degree of intraobserver and interobserver variance is possible; however, these large outlier values were uncommon.

Intraobserver agreement within 3 and 5 degrees was higher than interobserver agreement for all measurements (Tables 5 and 6). In light of its high interobserver reliability, interobserver agreement for the aMPFA was surprisingly the lowest for all frontal plane measurements.

DISCUSSION

We found good to very good interobserver reliability for all measurements except for the aADTA, which showed moderate reliability. Intraobserver reliability was slightly

TABLE 4. Median Absolute Difference and Range for Sagittal Plane Measurements

	Intraobserver (degrees)				Interobserver (degrees)
Measurement	**Observer 1**	**Observer 2**	**Observer 3**	**Mean**	
aPDFA (n = 46)	2.0 (0–9)	2.0 (0–14)	3.0 (0–18)	2.3	3.0 (0–18)
aPPTA (n = 64)	2.0 (0–10)	2.0 (0–10)	3.0 (0–22)	2.3	3.0 (0–18)
aADTA (n = 62)	2.0 (0–8)	3.0 (0–15)	3.0 (0–17)	2.7	3.0 (0–18)
Mean	2.0	2.3	3.0	2.4	3.0

TABLE 5. Percent Agreement Within 3 and 5 Degrees Between Frontal Plane Measurements

	Intraobserver (3/5)				Interobserver (3/5)
Measurement	**Observer 1**	**Observer 2**	**Observer 3**	**Mean**	
aMPFA (n = 60)	75.0/86.7	56.7/78.3	68.3/85.0	66.7/83.3	23.3/51.7
aLDFA (n = 60)	83.3/96.7	78.3/90.0	86.7/95.0	82.8/93.9	60.0/80.0
mMPTA (n = 60)	85.0/93.3	66.7/76.7	88.3/95.0	80.0/88.3	46.7/56.7
mLDTA (n = 60)	75.0/93.3	66.7/88.3	71.7/90.0	71.1/90.5	45.0/73.3
Mean	79.6/92.5	67.1/83.3	78.8/91.3	75.2/89.0	43.8/65.4

higher than interobserver reliability for all measurements and was also good to very good with 1 moderate value (also the aADTA). Frontal plane measurements showed higher reliability than sagittal plane measurements. Our findings indicate that these lower extremity measurements are a reliable method of assessing limb deformity; however, the aADTA should be interpreted with some skepticism and may warrant repeat measurement.

The merit of lower extremity deformity measurement is further evidenced by the values obtained for the median absolute differences. The greatest median absolute difference in this study was 3.0 degrees, indicating that greater than half of all measurements for a given angle on a given radiograph were within 3 degrees of one another.

Our analysis of agreement provides a surprising and seemingly contradictory finding. Although proximal femoral measurements had the highest interobserver reliability (0.91) and the second highest intraobserver reliability (0.93), these measurements have the lowest agreements within 3 and 5 degrees of any in the study (Tables 1 and 5). This apparent conflict is explained by noting that the range in values of the proximal femoral measurements is far less than those of the proximal and distal tibia, and the median value of the proximal femoral measurements is larger than the other values. Therefore, the proximal femoral measurements lie within a narrower range; however, within that range, there is greater heterogeneity of values.

As expected from the respective ICC (2,1) values, the ranges of values for measurements of the tibia tended to be greater than those of the femur. Interestingly, the greatest range in intraobserver measurements for all 3 observers occurred with the measurement of the mMPTA in the frontal plane. The mMPTA also had the highest range among the interobserver measurements. Therefore, despite good reliability, inherent in this measurement is the potential for wide-ranging variance; its reported value should be carefully interpreted, and repeat measurement may be prudent.

Several other investigations concerning lower extremity measurement reliability have been conducted.[6–11] As the raw data are not included in many of these articles and as their statistical analysis methods are not uniform, it is difficult to draw conclusions as to the comparable reliability of the measurements in the current study. However, Myers et al[8] reported intraobserver and interobserver ICC (version not specified) values for the aLDFA of 0.97 and 0.92, respectively, and for the mLDTA of 0.62 and 0.92, respectively. These values correlate with our findings (Tables 1 and 2).

Mechanical axis reproducibility was studied by Henderson and Kemp,[6] who reported 100% agreement within 3 degrees between 2 independent observers. The ICC values for mechanical axis angle and mechanical axis deviation were reported by Specogna et al.[11] When viewing plain films, intraobserver ICC values for mechanical axis angle and mechanical axis deviation averaged 0.98 and 0.99, respectively. Interobserver ICC values for mechanical axis angle and mechanical axis deviation were 0.96 and 0.99, respectively.

Odenbring et al[9] reported a mean absolute difference between hip-knee-ankle angle of 1.3 degrees, with an absolute difference range of 0 to 2 degrees.

All but one of the measurements in this study were found to have reliabilities of good or better, with only one considered moderate. Median absolute difference values were less than or equal to 3 degrees for all measurements. Although a small minority of measurements had a wide distribution of values, reliability and agreement of lower extremity deformity measurements were acceptable. Lower extremity deformity measurements should continue to be used in guiding treatment and determining outcome. The orthopaedic surgeons involved with deformity correction should become proficient in performing these measurements.

TABLE 6. Percent Agreement Within 3 and 5 Degrees Between Sagittal Plane Measurements

	Intraobserver (3/5)				Interobserver (3/5)
Measurement	**Observer 1**	**Observer 2**	**Observer 3**	**Mean**	
aPDFA (n = 46)	76.1/93.5	69.6/89.1	52.2/71.7	66.0/84.8	34.8/54.3
aPPTA (n = 64)	76.6/85.9	59.4/82.8	60.9/79.7	65.6/82.8	35.9/54.7
aADTA (n = 62)	75.8/91.9	53.2/80.6	56.5/75.8	61.8/82.8	43.5/67.7
Mean	76.2/90.4	60.7/84.1	56.5/75.7	64.4/83.4	38.1/58.9

ACKNOWLEDGMENTS

The authors thank Springer-Verlag for permission to include the illustration of standard lower extremity deformity measurements.

REFERENCES

1. Paley D, Herzenberg JE, Tetsworth K, et al. Deformity planning for frontal and sagittal plane corrective osteotomies. *Orthop Clin North Am*. 1994;25:425–465.
2. Paley D, Tetsworth K. Mechanical axis deviation of the lower limbs. Preoperative planning of multiapical frontal plane angular and bowing deformities of the femur and tibia. *Clin Orthop*. 1992;280:65–71.
3. Paley D, Tetsworth K. Mechanical axis deviation of the lower limbs. Preoperative planning of uniapical angular deformities of the tibia or femur. *Clin Orthop*. 1992;280:48–64.
4. Paley D. *Principles of Deformity Correction Exercise Workbook*. Berlin, Germany: Springer-Verlag; 2003.
5. Altman DG. *Practical Statistics for Medical Research*. New York, NY: Chapman and Hall; 1991.
6. Henderson RC, Kemp GJ. Assessment of the mechanical axis in adolescent tibia vara. *Orthopedics*. 1991;14:313–316.
7. Ilahi OA, Kadakia NR, Huo MH. Inter- and intraobserver variability of radiographic measurements of knee alignment. *J Knee Surg*. 2001;14:238–242.
8. Myers TG, Fishman MK, McCarthy JJ, et al. Incidence of distal femoral and distal tibial deformities in infantile and adolescent Blount disease. *J Pediatr Orthop*. 2005;25:215–218.
9. Odenbring S, Berggren A, Peil L. Roentgenographic assessment of the hip-knee-ankle axis in medial gonarthrosis. *Clin Orthop*. 1993;289:195–196.
10. Schmidt GL, Altman GT, Dougherty JT, et al. Reproducibility and reliability of the anatomic axis of the lower extremity. *J Knee Surg*. 2004;17:141–143.
11. Specogna AV, Birmingham TB, DaSilva JJ, et al. Reliability of lower limb frontal plane alignment measurements using plain radiographs and digitalized images. *J Knee Surg*. 2004;17:203–210.