Measurement Variance in Limb Length Discrepancy

Clinical and Radiographic Assessment of Interobserver and Intraobserver Variability

Michael A. Terry, MD, Jennifer J. Winell, MD,‡ Daniel W. Green, MD,† Robert Schneider, MD§ Margaret Peterson, MD,* Robert G. Marx, MD, MSc, FRCSC,¶ and Roger F. Widmann, MD†*

Abstract: The purpose of this study was to assess interobserver and intraobserver variability in the assessment of clinical and radiographic measurement of lower limb length discrepancy. Clinical measurements included direct measurement with a tape measure from anterior superior iliac spine (ASIS) to lateral malleolus and ASIS to medial malleolus as well as block measurement. Slit scanogram radiographic measurement was also evaluated. All three clinical measurements had excellent reliability, but the relatively large mean differences and the large 95% confidence intervals for clinical measurements limit the usefulness of these techniques. Slit scanogram measurement was the most reliable measurement technique. The intraobserver variance of direct slit scanogram measurement included intraclass correlation coefficient of 0.99, mean difference of 0.1 cm, and 95% confidence interval of 0.4 cm. Results were not influenced by patient age or body mass index. Slit scanogram measurement is the preferred method for assessment of limb length discrepancy. The direct slit scanogram measurement described in the text follows the mechanical axis line of the leg in the "at ease" standing position described by Paley. Direct measurement using a measuring tape on a full-length slit scanogram is more reliable than indirect measurement using horizontal lines drawn to a radiolucent ruler that is positioned by a technician, since direct measurement avoids errors due to nonparallel positioning of the limb relative to the ruler, and direct measurement also avoids errors due to non-horizontal lines drawn from standard bony landmarks to the ruler. The ideal radiographic measurement technique would have high reliability and accuracy and would minimize or eliminate radiation.

Key Words: scanogram, epiphysiodesis, limb length discrepancy

(*J Pediatr Orthop* 2005;25:197–201)

Study conducted at the Hospital for Special Surgery, New York, New York.
From *Hospital for Special Surgery, Weill Medical College of Cornell University, New York, New York; †Division of Pediatric Orthopaedic Surgery, Hospital for Special Surgery, Weill Medical College of Cornell University, New York, New York; ‡Division of Pediatric Orthopaedic Surgery, Montefiore Medical Center, Bronx, New York; §Department of Radiology, Hospital for Special Surgery, Weill Medical College of Cornell University, New York, New York; and ¶Foster Center for Clinical Outcome Research, Hospital for Special Surgery, Weill Medical College of Cornell University, New York, New York.
None of the authors received financial support for this study.
Reprints: Roger F. Widmann, MD, Hospital for Special Surgery, 535 East 70th Street, New York, NY 10021 (e-mail: widmannr@hss.edu).

Accurate prediction of limb length discrepancy (LLD) at maturity depends on both reliable acquisition of radiographic data and accurate prediction of future discrepancy.[24] Clinical assessment of LLD is notoriously unreliable, with block measurement widely believed to be the most reliable clinical measurement technique.[15,23] A variety of radiographic techniques have been developed, including the teleoroentgenogram,[23] orthoroentgenogram,[11] slit scanogram,[21,22,29] computed tomography (CT) scanogram,[1,2,12,14] and ultrasound.[16,33]

While the reliability of radiographic techniques, including orthoroentgenogram, CT scanogram, and ultrasound, has been assessed, the reliability of slit scanogram measurement has not been evaluated. Despite this fact, slit scanogram has been widely accepted as an accurate plain radiographic technique for limb length assessment. This technique uses a narrow slit x-ray beam that moves down the extremity and is orthogonal to the extremity at all times. This technique is thought to minimize parallax and magnification errors. A ruler is generally placed in the midline, and the lower limbs are positioned in neutral alignment parallel to the ruler.

There are several different methods available to calculate projected LLD at maturity based on past measurements of radiographic LLD.[3,25,27,34] Errors in the calculation of ultimate LLD may result from errors in leg length measurement, errors in assessment of skeletal age, errors in methodology for calculation of predicted LLD, physician math and documentation error, and inappropriate assumptions regarding normal and abnormal growth patterns.[4,11,31]

The outcomes of lower limb epiphysiodesis are quite variable.[4,10,18–20,32] A consistent percentage of patients fail to achieve limb lengths within 2 cm at maturity. Most studies attribute less-than-optimal results to variability in skeletal age assessment, faulty growth assumptions, or flaws in the methodology used to calculate the ultimate discrepancy. Variability in slit scanogram measurement has not been previously evaluated as a source of error in the prediction of LLD at maturity.

The purpose of this study was to assess the reliability of clinical measurements of LLD and the reliability of the slit scanogram measurement of LLD.

MATERIALS AND METHODS

Sixteen patients with documented lower limb discrepancies of various magnitudes and multiple etiologies were identified, and consent was obtained to allow them to participate in this IRB-approved study. The patients were chosen because they were all actively being followed for LLD, and all were scheduled for scanogram studies within a 2-month time frame of the study date. None of the patients had flexion contractures of the lower extremity. The patients represented a wide range of age, body habitus, and LLD (Table 1). Age range was 1.5 years to 19 years. Etiology of LLD included osteomyelitis, femur fracture, congenital pseudarthrosis, Blount disease, hemihypertrophy, fibular hemimelia, tibial hemimelia, and congenital short femur.

The observers were blinded to the patients' names, diagnoses, and history of prior treatment. Four observers (a pediatric orthopaedic surgeon, a pediatric orthopaedic fellow, a fourth-year orthopaedic resident, and a physical therapist) performed the clinical measurements of LLD. The clinical measurements included direct measurement of each limb, measuring from the top of the anterior superior iliac spine (ASIS) to the bottom of the lateral malleolus, and a separate set of measurements from the top of the ASIS to the bottom of the medial malleolus. Measurements of limb length were made using new standard woven nursing tape measures (Abco Dealers, Nashville, TN). A third clinical measurement of LLD was performed using blocks to level the pelvis as viewed and palpated from posterior. Sets of blocks measuring 0.25, 0.5, 1.0, 2.0, and 3.0 cm were machined from black Delrin plastic with tolerances of less than 0.01 cm (Integrated Machine and Engraving, Sussex, NJ). Patients and observers were instructed not to discuss the results of any measurements. All results were recorded in millimeters. LLD was calculated by computer using the total limb length measurements from the tape measure to avoid mathematical errors in calculation of discrepancy.

Each observer performed all three clinical measurements on each of the 16 patients, and each observer performed repeat clinical measurements on a subset of 4 patients to provide data for intraobserver variance determination. The repeat examination was performed after all of the other patients were measured to maximize the interval between measurements (approximately 4 hours). The order of the three clinical measurements on each patient and the order of the patients were predetermined randomly using the extra period Latin square design.[13] Measurements were recorded and collected on each patient prior to evaluation of the next patient.

Each of the 16 patients had a radiographic assessment of limb length using the supine slit scanogram technique.[21] The film focal distance is standardized at 40 inches at our institution. The patient is placed on top of a long radiolucent ruler on a long 36-inch cassette, and both lower limbs are placed in neutral alignment parallel to the midline ruler.

The three physician observers participated in the evaluation of the slit scanograms for LLD. The names on the slit scanograms were covered with tape, and each physician measured each leg length independently. The slit scanograms were measured four times using two measurement techniques. For the indirect slit scanogram measurement, the observer marked the slit scanogram at the most proximal aspect of the femoral head and the middle of the distal tibial plafond, extending a horizontal line from these landmarks to the midline radiolucent ruler. Identical protractors (United States Manufacturing Company, Pasadena, CA) and identical thin radiographic marking pencils (Dixon Ticonderoga Co, Heathrow, FL) were used to mark the slit scanograms. The proximal and distal measurements were recorded from the midline radiolucent ruler, and all pencil marks were then erased. Calculation of LLD was performed afterward by computer using the proximal and distal measurements; thus, there was no opportunity for the observer to introduce mathematical errors into the calculation of discrepancy. The direct slit scanogram measurement

TABLE 1. Demographics of Study Population

Patient	Age	Height (cm)	Weight (kg)	BMI	Mean LLD	Etiology of LLD
1	16	152.5	44.4	19.09	8.88	Osteomyelits
2	4	118	23.1	16.59	9.63	Congenital short pemur
3	9	129	43	25.84	4.73	Hemihypertrophy
4	18	174.5	71.1	23.35	4.94	Congenital pseudarthrosis tibia
5	5	105.5	16.5	14.82	1.81	Hemihypertrophy
6	4	100.5	14.9	14.75	3.58	Osteomyelits
7	14	156.5	97.1	39.65	2.04	Blount Dx
8	4	162.5	63	23.86	2.46	Femur fracture
9	15	191	133	36.46	3.71	Blount Dx
10	1.5	85	10.4	14.39	0.30	Hemihypertrophy
11	19	168	60.5	21.44	4.23	Osteomyelits
12	10	141	37.1	18.66	3.63	Fibular hemimelia
13	10	127	21.2	13.14	4.63	Osteomyelits
14	3	95.5	15.7	17.21	2.15	Tibial hemimelia
15	14	178.5	97.8	30.69	5.23	Blount Dx
16	13	155	97.8	40.71	4.52	Fibular hemimelia
Mean	9.97	140.0	52.9	23.17	4.15	
Min	1.5	85	10.4	13.14	0.30	
Max	19	191	133	40.71	9.63	

was performed using a tape measure to directly measure leg length from a mark on the most proximal aspect of the femoral head to a mark in the middle of the distal tibial plafond (ignoring the radiolucent ruler).

The interobserver and intraobserver reliabilities of the three clinical and the two radiographic measurements of LLD were assessed by calculating the intraclass correlation coefficients (ICC),[30] the mean absolute differences,[5–7] and the 95% confidence intervals within and between examiners. The ICC is a statistical measure analogous to the kappa statistic used to evaluate agreement where 1.0 is perfect agreement. The mean absolute difference is the mean difference between two measures irrespective of the direction of the difference. The 95% confidence intervals were based on the observed distribution, with 95% of all measures falling between these intervals. The accuracy of the various techniques was not assessed, since a direct caliper measurement or other gold standard is not available for comparison.

RESULTS

All four observers performed three clinical measurements of LLD on each of the 16 patients, and each examiner performed a repeat examination of 4 patients. Each of the three physicians also evaluated each of the 16 slit scanograms for LLD using indirect measurements twice and direct measurements twice.

To determine intraobserver variability, we examined differences between two clinical or radiographic measurements of LLD made by the same observer. Reliability of clinical and radiographic measurement of LLD was assessed using the ICC. The ICCs for the clinical measurements by the same examiner (intraobserver reliability) were 0.88, 0.78, and 0.86 for measurement of LLD from ASIS to lateral malleolus and ASIS to medial malleolus and for blocks. The ICC for intraobserver reliability of indirect slit scanogram measurement was 0.94; for direct slit scanogram measurement it was 0.99. Table 2 summarizes the data we obtained for intraobserver variance, including mean absolute differences and 95% confidence intervals. We did not calculate 95% confidence intervals for the three clinical measurements since there were too few repeat clinical evaluations to yield statistically meaningful data.

The data for interobserver variance of clinical and radiographic measurements are summarized in Table 3. Once again, the direct slit scanogram measurement was the most reliable technique for assessment of limb length.

We performed two additional analyses to assess the impact of patient age and body mass index (BMI) on the reliability of clinical and radiographic measurement techniques. Neither age nor BMI had a significant effect on the reliability of the data.

TABLE 2. Intraobserver Variance: Summary of Clinical and Radiographic Measurements

Measurement	ICC	MD (cm)	95% CI
ASIS to lateral malleolus	0.88	0.89	NC
ASIS to medial malleolus	0.78	1.07	NC
Blocks	0.86	0.81	NC
Indirect slit scanogram	0.94	0.58	1.8
Direct slit scanogram	0.99	0.13	0.4

ICC, intraclass correlation coefficient; MD, mean difference; CI, confidence interval; NC, not calculated.

TABLE 3. Interobserver Variance: Summary of Clinical and Radiographic Measurements

Measurement	ICC	MD (cm)	95% CI
ASIS to lateral malleolus	0.83	1.11	2.6
ASIS to medial malleolus	0.8	1.03	2.5
Blocks	0.83	1.01	2.2
Indirect slit scanogram	0.89	0.7	1.7
Direct slit scanogram	0.98	0.28	1.1

ICC, intraclass correlation coefficient; MD, mean difference; CI, confidence interval.

DISCUSSION

The use of blocks under the short leg to level the pelvis has been shown to be the most reliable clinical test for LLD.[35] Jonson and Gross examined intraobserver and interobserver variability in LLD assessment using blocks to level the pelvis.[15] That study of 18 Naval Reserve officers noted intraobserver and interobserver ICCs for the block method of 0.87 and 0.70, respectively.

Our study results suggest that the interobserver reliability of all three methods of clinical LLD assessment (ASIS to medial malleolus, ICC = 0.83; ASIS to lateral malleolus, ICC = 0.80; block measurement, ICC = 0.83) was similar. However, the 95% confidence interval was smaller for block measurements compared with the other clinical measurements using a tape measure from the ASIS (see Table 3). The 95% confidence interval for interobserver measurement of LLD using blocks measured 2.2 cm in our study. This large 95% confidence interval is unacceptable for clinical decision making, and these results confirm the observation by Green et al that clinical measurement of LLD may be grossly inaccurate.[11]

There are several methods available for the radiographic assessment of LLD, such as teleoroentgenogram,[23] orthoroentgenogram,[11,14] slit scanogram,[21,22,29] ultrasound,[33] and CT scanogram.[1,2,12,14] The reliability of orthoroentgenograms and CT scanograms has been extensively studied, and in the absence of flexion contractures they have similar reliability and accuracy.[1,2,14,28] Some of the advantages of CT scanogram over orthoroentgenogram shown by these studies are decreased radiation and better assessment of limb length in the setting of flexion deformity. To our knowledge the reliability of the orthoroentgenogram and the CT scanogram has not been compared with the slit scanogram technique. The interobserver ICC for CT scanogram has been shown to be as high as 0.99,[2] which is quite similar to the interobserver ICC for direct measurement of slit scanogram (0.98) generated by this study.

Although the slit scanogram radiograph remains the preferred method for radiographic assessment of LLD in many centers, including our own, the reliability of slit scanogram radiographs has not been specifically studied previously. The advantages of the slit scanogram include avoiding magnification errors and potentially decreasing the size of the x-ray film.[23] We measured the radiographic LLD from the slit scanogram using two different techniques. The indirect slit scanogram measurement method has been used historically at our institution. We also used a direct radiographic slit scanogram measurement, since we hypothesized that this may be a more reliable technique.

Direct measurement of limb length with a tape measure from the slit scanogram proved more reliable than indirect measurement with a midline radiolucent ruler. Direct slit scanogram measurement had an intraobserver ICC of 0.99 and an interobserver ICC of 0.98. We no longer use indirect measurement of slit scanograms. Significant sources of error associated with indirect measurement include abduction or adduction of the limb relative to the midline ruler[21] and non-horizontal lines from bony landmarks to the midline ruler.

Accurate prediction of LLD at maturity depends on reliable and accurate assessment of LLD. Several excellent retrospective studies have addressed the issue of "failure" of epiphysiodesis to achieve equal limb length at maturity[4,10,18–20,32] (Table 4). With failure defined as LLD at maturity of greater than 2 cm, the average failure rate in several large (>50 patients) retrospective studies averaged 23% (range 7–40%). Blair et al[4] give a comprehensive discussion and explanation for the high failure rates noted historically; some of the reasons for failure, excluding surgical technique, were errors in timing and calculation of predicted LLD, error in estimation of maturity, and physician documentation errors. Reliability of radiographic limb length measurement is not addressed in any of these studies.

Little et al compared three methods for calculating LLD at maturity (the Anderson, Menelaus and Moseley techniques) and found no meaningful effect on the accuracy of predictions at maturity.[19] The multiplier method was recently described, and the results of this method correlate well with the Moseley method.[27] Although skeletal maturity assessment was thought to increase the accuracy of LLD predictions,[3] this has been called into question in clinical practice.[17,19] In addition, skeletal age determination has been shown to have large variability.[8,9]

The relatively high ICCs for the three clinical measurements and for the indirect slit scanogram radiographic measurement technique in this study suggest that these techniques might be clinically useful. However, the large mean absolute differences and the large 95% confidence intervals for all of these measurements are unacceptable in clinical practice. The direct slit scanogram measurement technique is extremely reliable and is the preferred method for limb length measurement at our institution.

The direct slit scanogram measurement follows the mechanical axis line of the leg, and the mechanical axis line is perpendicular to the ground in the "at ease" standing position described by Paley.[26] Thus, direct measurement is a functional assessment of limb length. Direct measurement also avoids errors due to nonparallel positioning of the limb relative to the axis of the x-ray film and the radiolucent ruler, as well as errors due to non-horizontal lines drawn from bony landmarks to the ruler. The ideal radiographic measurement technique would have high reliability and accuracy and would minimize or eliminate radiation.[23]

TABLE 4. Historical Failure Rate of Epiphysiodesis

Author	Year	% ≥2 cm	n
Green Anderson	1956	10.4	125
Stamp-Lansche	1960	40	104
Menelaus	1966	7	94
Blair	1982	34	67
Little	1996	27	71
Kemnitz	2003	17.5	57

Failure rate defined as percentage of limb length discrepancies at maturity ≥2 cm.

ACKNOWLEDGMENTS

The authors thank John Denneen, PT, and Karin Wilk, Coordinator, Division of Pediatric Orthopaedics, for their assistance and significant efforts in organizing and carrying out this study.

REFERENCES

1. Aaron A, Weinstein D, Thickman D, et al. Comparison of orthoroentgenography and computed tomography in the measurement of limb-length discrepancy. *J Bone Joint Surg [Am]*. 1992;4:897–902.
2. Aitken AG, Flodmark O, Newman DE, et al. Leg length determination by CT digital radiography. *AJR Am J Roentgenol*. 1985;144:613–615.
3. Anderson M, Green W, Messner M. Growth and predictions of growth in the lower extremities. *J Bone Joint Surg [Am]*. 1963;45:1–14.
4. Blair VP, Walker SJ, Sheridan JJ, et al. Epiphysiodesis: a problem of timing. *J Pediatr Orthop*. 1982;2:281–284.
5. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*. 1995; 346:1085–1087.
6. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. *Comput Biol Med*. 1990;20:337–340.
7. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
8. Carpenter CT, Lester EL. Skeletal age determination in young children: analysis of three regions of the hand/wrist film. *J Pediatr Orthop*. 1993; 13:76–79.
9. Cundy P, Paterson D, Morris L, et al. Skeletal age estimation in leg length discrepancy. *J Pediatr Orthop*. 1988;8:513–515.
10. Green WT, Anderson M. Epiphyseal arrest for the correction of discrepancies in length of the lower extremities. *J Bone Joint Surg [Am]*. 1957;39:853–872.
11. Green WT, Wyatt GM, Anderson M. Orthoroentgenography as a method of measuring the bones of the lower extremity. *J Bone Joint Surg [Am]*. 1946;28:60.
12. Helms CA, McCarthy S. CT scanograms for measuring leg length discrepancy. *Radiology*. 1984;151:802.
13. Hoaglin DC, Mosteller F, Tukey JW. *Fundamentals of Exploratory Analysis of Variance*. New York: John Wiley & Sons, Inc. 1991.
14. Huurman WW, Jacobsen FS, Anderson JC, et al. Limb-length discrepancy measured with computerized axial tomographic equipment. *J Bone Joint Surg [Am]*. 1987;69:699–705.
15. Jonson SR, Gross MT. Intraexaminer reliability, interexaminer reliability, and mean values for nine lower extremity skeletal measures in healthy naval midshipmen. *J Orthop Sports Phys Ther*. 1997;25:253–263.
16. Junk S, Terjesen T, Rossvoll I, et al. Leg length inequality measured by ultrasound and clinical methods. *Eur J Radiol*. 1992;14:185–188.

17. Kasser JR, Jenkins R. Accuracy of leg length prediction in children younger than 10 years of age. *Clin Orthop*. 1997;338:9–13.
18. Kemnitz S, Moens P, Fabry G. Percutaneous epiphysiodesis for leg length discrepancy. *J Pediatr Orthop B*. 2003;12:69–71.
19. Little DG, Nigo L, Aiona MD. Deficiencies of current methods for the timing of epiphysiodesis. *J Pediatr Orthop*. 1996;16:173–179.
20. Menelaus MB. Correction of leg length discrepancy by epiphysial arrest. *J Bone Joint Surg [Br]*. 1966;48:336–339.
21. Merrill V. *Atlas of Roentgenographic Positions*. St Louis: CV Mosby, 1967.
22. Millwee RH. Slit scanography. *Radiology.* 1937;28:483–486.
23. Moseley CF. Leg length discrepancy. In: Morrisy RT, Weinstein SL, eds: *Lovell and Winter's Pediatric Orthopaedics*. New York: Lippincott Williams & Wilkins, 2001:1105–1150.
24. Moseley CF. A straight line graph for leg length discrepancies. *Clin Orthop*. 1978;136:33–40.
25. Moseley CF. A straight-line graph for leg-length discrepancies. *J Bone Joint Surg [Am]*. 1977;59:174–179.
26. Paley D. Normal lower limb alignment and joint orientation. In: *Principles of Deformity Correction*. New York: Springer; 2002:1–18.
27. Paley D, Bhave A, Herzenberg JE, et al. Multiplier method for predicting limb-length discrepancy. *J Bone Joint Surg [Am]*. 2000;82:1432–1446.
28. Porat S, Fields S. Limb length discrepancy determined by computerized tomography and radiography. *Harefuah*. 1989;116:515–516.
29. Pugh DG, Winkler NT. Scanography for leg-length measurement: an easy satisfactory method. *Radiology.* 1966;87:130–133.
30. Rosner B. *Fundamentals of Biostatistics*. Toronto: Duxbury Press, 1995.
31. Shapiro F. Developmental patterns in lower-extremity length discrepancies. *J Bone Joint Surg [Am]*. 1982;64:639–651.
32. Stamp WG, Lansche WE. Treatment of discrepancy in leg length. *South Med J.* 1960;53:764–774.
33. Terjesen T, Benum P, Rossvoll I, et al. Leg-length discrepancy measured by ultrasonography. *Acta Orthop Scand*. 1991;62:121–124.
34. Westh RN, Menelaus MB. A simple calculation for the timing of epiphysial arrest: a further report. *J Bone Joint Surg [Br]*. 1981;63:117–119.
35. Woerman AL, Binder-Macleod SA. Leg length discrepancy assessment: accuracy and precision in five clinical methods of evaluation. *J Orthop Sports Phys Ther.* 1984;5:230–239.