Reliability Analysis for Radiographic Measurement of Limb Length Discrepancy

Full-Length Standing Anteroposterior Radiograph Versus Scanogram

Sanjeev Sabharwal, MD, Caixia Zhao, MD,* John McKeon, MD,* Todd Melaghari, MD,* Marcia Blacksin, MD,† and Cornelia Wenekor, MD†*

Abstract: Patients with limb length discrepancy (LLD) often have associated angular deformities requiring a standing full-length radiograph of the lower limb in addition to a scanogram. The purpose of our study was to determine the intraobserver and interobserver reliability of measuring LLD with both techniques, using computed radiography. The LLD was measured on 70 supine scanograms and standing anteroposterior radiographs of the lower extremity by 5 blinded observers on 2 separate occasions. Intraclass correlation coefficient (ICC) and mean absolute difference (in millimeters) was calculated to assess intraobserver and interobserver reliability and found to be excellent for both radiographic techniques. Intraobserver ICC and mean absolute difference was 0.975 to 0.995 and 1.5 to 2.6 mm for scanogram and 0.939 to 0.996 and 1.5 to 4.6 mm for the standing radiograph, respectively. Repeated measurements for both radiographic studies were within 5 mm of the first measurement greater than 90% and within 10 mm greater than 95% of times. Interobserver ICC and mean absolute difference was 0.979 and 2.6 mm for scanogram and 0.968 and 3.0 mm for the standing radiograph. The reliability was excellent irrespective of age, sex, and underlying diagnosis other than Blount disease, which had good reliability. A standing anteroposterior radiograph of the lower extremity should be the imaging modality of choice when evaluating patients with limb length inequality who may have angular deformities because it allows a comprehensive evaluation of the extremity and is as reliable as a scanogram for measuring LLD. This approach may decrease the radiation exposure and financial burden involved in assessing patients with unequal limb lengths.

Key Words: limb length discrepancy, reliability, radiographic analysis, computed radiography

(*J Pediatr Orthop* 2007;27:46–50)

Limb length discrepancy (LLD) is a relatively common diagnosis[1] in patients with spine and lower extremity disorders. Significant LLD can cause gait deviations and lead to degenerative changes in the joints of the lower extremity and lumbar spine.[1–3] The diagnosis of LLD is usually made by history and physical examination and confirmed with imaging studies. A large percentage of patients with LLD may also have angular deformities of the femur and/or tibia that require a standing anteroposterior (AP) radiograph of the entire lower extremity for deformity analysis. Before making treatment recommendations, it is imperative for the clinician to have a reliable imaging modality to quantify LLD and commonly associated limb deformities.[2–4]

Some authors[4,5] have questioned the need for a scanogram for routine radiographic analysis of LLD and proposed that a standing full-length AP radiograph of the lower extremity is a more comprehensive study because it cannot only analyze limb deformities but can also determine LLD. We currently use a standing full-length AP radiograph of the lower extremity (modified teleoroentgenogram)[2] for deformity analysis and a scanogram with a radiopaque ruler for assessment of LLD. Both radiographs are done using computed radiography (CR), which is distinct from computerized tomography (CT). In CR, a latent image is produced that is stored on a photostimulable phosphor receptor contained in a standard radiographic cassette. The image is transferred digitally and manipulated by an automated system such as the picture archiving and communications system (PACS), resulting in a film radiograph. The operator can enhance the final image by using the computer to adjust the image parameters. As a result, quality radiographs can be obtained consistently with a significant reduction in the radiation dose compared with standard film-screen systems.[6] This technique may be particularly useful in children and young adults who require repeated radiographic examinations because of limb length inequality.

The purpose of our study was to determine the intraobserver and interobserver reliability of measuring LLD using a scanogram and full-length standing AP radiograph of the lower extremity using CR.

METHODS

After approval from the institutional review board, a retrospective review of patients who had a standing full-length AP radiograph of the lower extremity and a scanogram using CR performed on the same day over a 3-year period was carried out. Patients were excluded if they had an overlying external fixator, the hip or ankle joint was not visualized on either of the imaging studies or if the scanogram ruler did not extend to the level of the hip or ankle joint. Of the 114 patients who met the study criteria, 70 patients were randomly selected for this study (power = 0.94).

From the Departments of *Orthopedics and †Radiology, New Jersey Medical School, Newark, NJ.
None of the authors received financial support for this study.
Reprints: Sanjeev Sabharwal, MD, Associate Professor, Department of Orthopaedics, New Jersey Medical School, 90 Bergen Street, DOC 7300, Newark, NJ 07103. E-mail: sabharsa@umdnj.edu.

The standing AP radiograph was performed with the patient facing the radiographic tube and both patellas pointing forward (Fig. 1). An attempt was made to level the pelvis with an appropriate-sized lift placed under the short limb. The minimum patient to tube distance was 203 cm and was increased for taller individuals. The radiograph was obtained using 70 to 80 kVp and 30 to 160 mAs, based on the patient's size. The images were recorded on a Kodak (Rochester, NY) DirectView CR Long-length Imaging System using a vertical cassette holder with 3 individual 35 × 43 cm CR storage phosphor cassettes.[7] The 3 images were then stitched to form a composite image that was distributed to the PACS workstation (Centricity PACS 2.0, GE Medical Systems Information Technologies, Milwaukee, Wis).

There is some inconsistency in the literature regarding the term "scanogram." The term may have been derived from the technique of slit scanography,[8] in which the x-ray beam is tightly collimated to a thin transverse slit that exposes the film as the x-ray tube is moved from one end of the limb to the other. Others[1,2,4] have used the word "scanogram" to describe a modification of the orthoroentgenogram taken with 3 separate exposures centered at the hip, knee, and ankle using a standard sized (35 × 43 cm) cassette as opposed to the long cassette (35 × 110 cm), as was originally described by Green et al.[9] In the current study, the scanogram was performed with the patient supine using CR. The lower limbs were positioned with both patella pointing toward the ceiling and a radiopaque ruler taped to the table between the limbs. The patient to tube distance was 101 cm. Three separate AP images were obtained centered over the hip, knee, and ankle joints using 3 separate 35 × 43 cm CR storage phosphor cassettes. Pelvis films were obtained using 80 kVp, 40 to 200 mAs; knees with 70 kVp, 10 to 20 mAs; and ankles using 60 kVp, 4 mAs.The images were then distributed to the PACS workstation (Fig. 2).

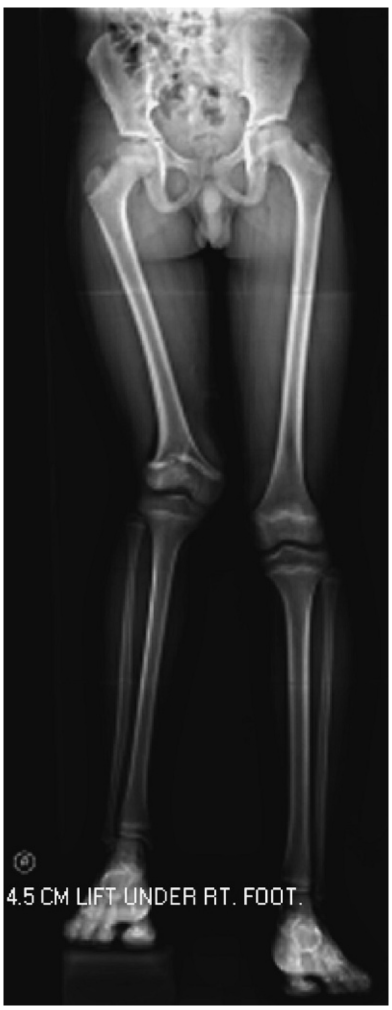

FIGURE 1. Standing AP radiograph of the lower extremities with the patient standing on a 4.5-cm lift under the short leg.

For both imaging studies, length of the lower limb was measured from the proximal end of femoral head to the center of the tibial plafond on each side and the difference (LLD) calculated in millimeters. Five blinded observers including 2 fellowship-trained pediatric orthopaedists, 2 musculoskeletal radiologists, and a senior (PGY 5) orthopaedic resident made all measurements on two separate occasions. Previous measurements were not available to any of the observers.

Statistical analysis was performed using SAS, version 9.1 software (SAS Institute Inc, Cary, NC). Analysis of variance was performed to estimate the intraobserver and interobserver reliability using intraclass correlation coefficients (ICC). An ICC of 0 to 0.24 reflects poor; 0.25 to 0.49, low; 0.50 to 0.69, fair; 0.7 to 0.89, good; and 0.9 to 1.0, excellent correlation.[10,11] Mean absolute difference in millimeters and 95% confidence intervals for measurements of LLD were also calculated for both imaging studies.[12] Effect of different variables such as patient's age, sex, and underlying diagnosis on the reliability of measuring LLD was assessed using analysis of variance.

RESULTS

There were 39 male (56%) and 31 female (44%) patients with an age range from 9 months to 73 years (mean, 20 years). Of these, 46 (66%) were younger than and 24 (34%) older than 18 years of age. The cause for LLD was posttraumatic in 42 (60%), congenital in 13 (19%), Blount disease in 9 (13%), and miscellaneous causes in the remaining 6 (8%) patients.

The magnitude of LLD ranged from 1 to 96 mm (mean, 23 mm) based on scanogram measurements and from 0 to 98 mm (mean, 21 mm) on standing AP radiograph. The intraobserver ICC and mean absolute difference among the 5 observers ranged from 0.975 to 0.995 and 1.5 to 2.6 mm for scanogram and 0.939 to 0.996 and 1.5 to 4.6 mm for standing radiograph (Table 1). The intraobserver agreement was within 5 mm of the original measurement 92% (range, 83%–97%) of the time when using a scanogram and 91% (range, 74%–100%) of the time when using the standing AP radiograph. The intraobserver agreement was within 10 mm of the original measurement in 98% (range, 97%–100%) of the patients when using a scanogram and in 96% (range,

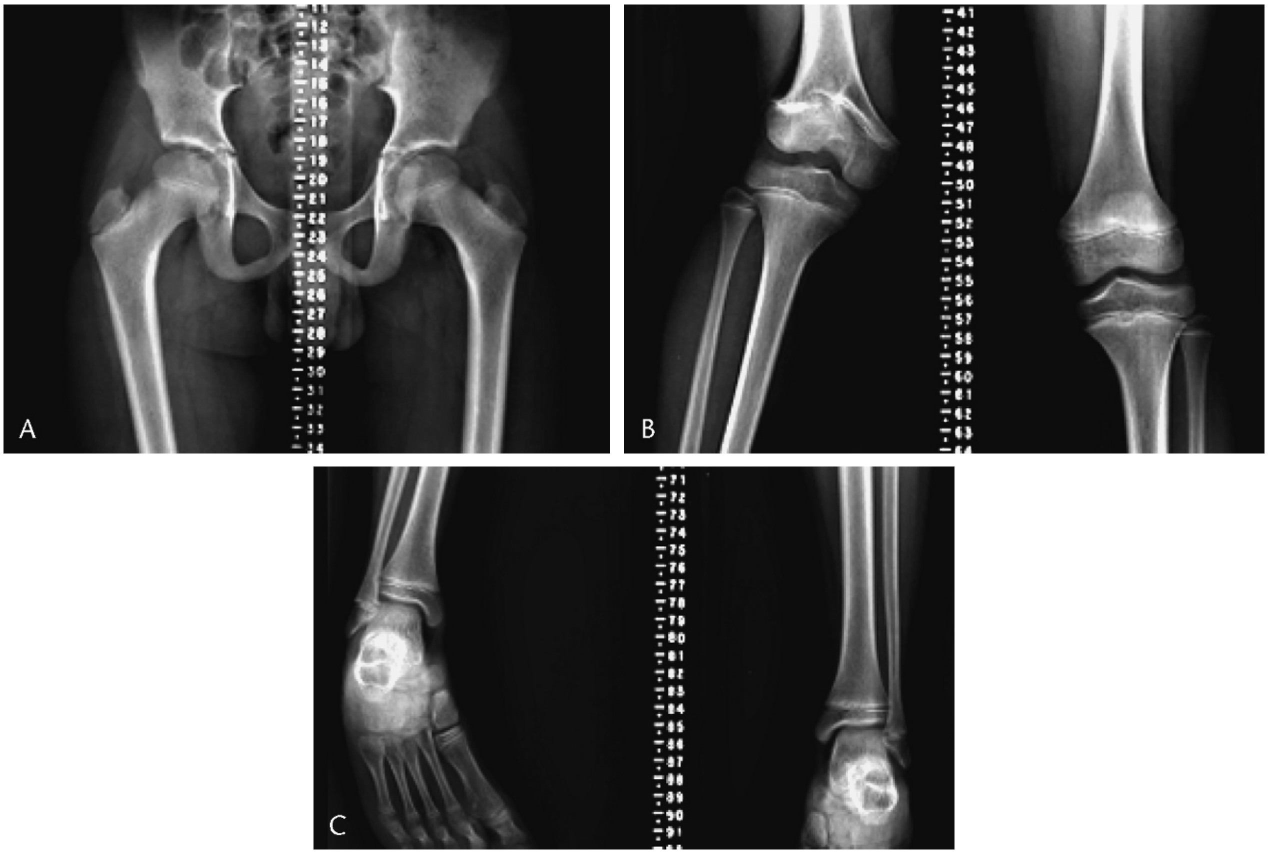

FIGURE 2. Scanogram of the same patient shown in Figure 1 with images centered at the hip (A), knee (B) and ankle (C) joints.

89%–100%) of the patients when using the standing AP radiograph.

The interobserver ICC and mean absolute difference was 0.979 and 2.6 mm for scanogram and 0.968 and 3.0 mm for standing radiograph (Table 2). The interobserver reliability was excellent irrespective of age, sex, and underlying diagnosis other than Blount disease, which had good reliability (ICC, 0.87 for scanogram and 0.77 for standing radiograph).

DISCUSSION

There is general agreement that radiographs are more accurate and reliable than clinical examination for analysis of LLD.[13–15] Accuracy refers to variation in measurement using a technique compared with actual length of the bone, whereas reliability is the variation between observers and within a single observer in obtaining the measurement.[16] One needs to consider several issues such as reliability, accuracy, magnification, radiation dose, cost, need for special equipment, convenience, and opportunity to image the entire extremity for etiology and angular deformities when choosing the imaging technique for evaluating patients presenting with LLD.

The issue of accuracy and magnification with the standing radiograph compared with the scanogram was not addressed in this study but has been previously reported. Sabharwal et al[5] reported on 111 patients with LLD who had a full-length standing AP radiograph and a scanogram performed on the same day using CR. Despite an average magnification of 4.6% (3.3 cm) when measuring lower extremity length using full-length standing radiograph, the mean difference in LLD measurement between the 2 techniques was only 0.5 cm. Green et al[9] compared the magnification produced by a teleoroentgenogram (a technique quite similar to the standing full-length radiograph used in the current study) with a supine orthoroentgenogram (similar to a scanogram in the current study). Using 10 adult human skeletons, they found a mean magnification of 4.5% (1.8 cm) for the femoral segment and 3% (0.9 cm) for the tibial segment. Because these specimens did not have LLD, the authors were unable to comment on the difference in limb lengths measured by the 2 imaging techniques. However, based on their clinical experience, they did comment that although the teleoroentgenogram may not measure the true length of the bone, it "fairly accurately" assesses the relative lengths of the 2 extremities at a single examination.

Although a scanogram may improve measurement accuracy by decreasing magnification and directly imaging at the joints, this is achieved at the expense of inability to visualize the entire length of the femur and tibia. Often, patients with LLD also have angular deformities and limb

TABLE 1. Intraobserver Reliability for Measuring LLD Using Scanogram and Standing AP Radiograph

	Orthopaedist 1	Orthopaedist 2	Radiologist 1	Radiologist 2	Resident
Scanogram					
ICC (95% CI)	0.995 (0.993–0.997)	0.985 (0.976–0.991)	0.989 (0.982–0.993)	0.994 (0.991–0.997)	0.975 (0.969–0.984)
MAD (mm) (95% CI)	1.6 (1.2–1.9)	2.6 (1.9–3.3)	2.0 (1.3–2.7)	1.5 (1.0–1.9)	2.0 (0.9–3.2)
Repeated measurement within 5 mm (%)	97	83	90	96	94
Repeated measurement within 10 mm (%)	100	97	97	99	97
Standing AP					
ICC (95% CI)	0.980 (0.969–0.988)	0.994 (0.991–0.996)	0.974 (0.959–0.984)	0.996 (0.994–0.998)	0.939 (0.904–0.962)
MAD (mm) (95% CI)	1.8 (0.8–2.9)	1.7 (1.2–2.1)	3.3 (2.3–4.3)	1.5 (1.2–1.8)	4.6 (0.3–6.3)
Repeated measurement within 5 mm (%)	97	96	87	100	74
Repeated measurement within 10 mm (%)	97	99	93	100	89

MAD indicates mean absolute difference in millimeters; 95% CI, 95% confidence interval.

malalignment, which may need to be addressed as part of the complete treatment strategy. Unlike the scanogram, a standing AP radiograph of the lower extremity incorporates the entire length of the limb including the foot height and allows for an in-depth analysis of the location and magnitude of frontal plane malalignment.[3] With visualization of the entire length of the femur and tibia, a more comprehensive analysis of the extremity can be performed including visualization of diaphyseal deformities, joint space narrowing, and instability. Moreover, because the weight-bearing radiograph involves a single radiographic exposure compared with 3 with the scanogram, there is a reduced risk of measurement error because of patient motion between radiographic exposures when using the standing radiograph. In addition, the accuracy of measuring LLD using a scanogram becomes difficult to maintain in patients with a large limb length discrepancy because the knee and ankle joints may be at significantly different levels on the 2 sides. In such cases, the contralateral knee or ankle joint may not be visualized on the limited AP view of the extremity afforded by the scanogram.

In a recent study, Leitzes et al[17] reported on the reliability of magnetic resonance imaging scanogram compared with slit scanogram and CT scanogram for measuring LLD by 3 blinded observers having varying levels of orthopaedic training. This was a cadaveric study where only the femoral segment was measured, and none of the "subjects" had any documented LLD. Despite these differences, the reported reliability of all 3 imaging techniques for measuring limb length was comparable with our clinical findings using other radiographic techniques. In the current clinical study, we assessed the reliability of 2 different imaging modalities that are commonly used at our institution for assessing patients with LLD, a supine scanogram and standing AP radiograph of the lower extremity. Using CR, both techniques demonstrated very high intraobserver and interobserver reliability, irrespective of patient's age, sex, and underlying diagnosis. The only variable that did not show excellent, but good reliability was a diagnosis of Blount disease. This may be related to the large size of typical patients with Blount disease combined with multiplanar limb deformities that can compromise adequate visualization of bony landmarks and can thus lead to a less reproducible measurement of limb length discrepancy on both types of radiographs.

TABLE 2. Interobserver Reliability for Measuring LLD Using Scanogram and Standing AP Radiograph

	ICC (95% CI)	MAD (mm) (95% CI)
Scanogram	0.979 (0.970–0.986)	2.6 (2.0–3.1)
Standing AP	0.968 (0.955–0.978)	3.0 (2.2–3.8)

MAD indicates mean absolute difference in millimeters; 95% CI, 95% confidence interval.

A large number of patients with significant LLD are young and may require several radiographs at frequent intervals to monitor limb length discrepancy and malalignment. Despite the use of new technologies such as CR and PACS, the issue of radiation exposure needs careful consideration. Although not the sole criteria for choosing the imaging modality of choice, the issue of cost needs to be considered as well. At our institution, the radiology charges (year 2005) are the same for both the scanogram and standing AP radiograph at US $137 each (technical component, $55; professional component, $82). If one was to assume that the scanogram did not provide any additional information, a significant cost saving could be accomplished with more prudent use of such imaging modalities.

Digital radiography and PACS technology is gaining popularity. Computer-aided and digital measurements do offer certain advantages such as decreased error associated with line drawing and manual rulers, ability to manipulate the magnification and contrast of an underpenetrated or overpenetrated film as well as improved storage and easier viewing capabilities using newer technology.[11] The improved reliability of digital compared with plain radiographs has been previously reported for spinal deformities[11,18] and distal radius fractures.[19] Although we did not compare plain radiographs with digital imaging in our study, we did find very acceptable reliability using CR for measuring LLD. Whether these findings can be duplicated in routine clinical practice with varied radiographic techniques and different observers will need further investigation.

In conclusion, information obtained from a variety of clinical and radiographic methods is often required before recommending treatment for patients with LLD. The choice of imaging modality should allow for a comprehensive analysis of the entire extremity in a safe, reliable, and cost-effective manner. Both the standing AP radiograph and the supine scanogram were found to have comparable and very high reliability for measuring LLD for a diverse group of patients. A standing AP radiograph of the lower extremities should be the imaging modality of choice when evaluating patients with limb length inequality who may have also have angular deformities because it allows for a more comprehensive evaluation of the lower extremity and is as reliable as a scanogram for measuring LLD. This approach may decrease the radiation exposure and financial burden involved in assessing such patients.

ACKNOWLEDGMENT

The authors thank Emily McClemens, PA-C, for her assistance in preparation of this manuscript.

REFERENCES

1. Shapiro F. *Pediatric Orthopedic Deformities. Basic Science, Diagnosis, and Treatment.* San Diego: Academic Press; 2001.
2. Moseley CF. Leg length discrepancy. In: Morrissy RT, Weinstein SL, eds. *Lowell and Winter's Pediatric Orthopaedics*. Philadelphia, PA: Lippincott Williams and Wilkins; 2001:1106–1145.
3. Paley D. *Principles of Deformity of Correction.* Berlin: Springer-Verlag Berlin Heidelberg; 2002.
4. Machen MS, Stevens PM. Should full-length standing anteroposterior radiographs replace the scanogram for measurement of limb length discrepancy? *J Pediatr Orthop B*. 2005;14:30–37.
5. Sabharwal S, Zhao C, McKeon J, et al. Computed radiographic measurement of limb length discrepancy—full length standing antero-posterior radiograph versus scanogram. *J Bone Joint Surg [Am]* 2006;88:2243–2251.
6. Kogutt MS. Computed radiographic imaging: use in low-dose leg length radiography. *AJR Am J Roentgenol*. 1987;148:1205–1206.
7. Wang X, Food DH, Doran J, et al. Fully automatic and reference-marker-free image stitching method for full-spine and full-leg imaging with computed radiography. *Proc. SPIE*. 2004;5368:361–369.
8. Millwee RH. Slit scanography. *Radiography*. 1937;28:483–486.
9. Green WT, Wyatt GM, Anderson M. Orthoroentgenography as a method of measuring the bones of the lower extremities. *J Bone Joint Surg [Am]*. 1946;28:60–65.
10. Munro BH. Correlation. In: Munro BH, ed. *Statistical Methods for Health Care Research*, 5th ed. Philadelphia, PA: Lippincott-Raven; 2005:239–258.
11. Kuklo TR, Potter BK, O'Brien MF, et al. Reliability analysis for digital adolescent idiopathic scoliosis measurements. *J Spinal Disord Tech*. 2005;18:152–159.
12. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307–310.
13. Terry MA, Winell JJ, Green DW, et al. Measurement variance in limb length discrepancy: clinical and radiographic assessment of interobserver and intraobserver variability. *J Pediatr Orthop*. 2005;25:197–201.
14. Cleveland RH, Kushner DC, Ogden MC, et al. Determination of leg length discrepancy. A comparison of weight-bearing and supine imaging. *Invest Radiol*. 1988;23:301–304.
15. Lampe HI, Swierstra BA, Diepstraten AF. Measurement of limb length inequality. Comparison of clinical methods with orthoradiography in 190 children. *Acta Orthop Scand*. 1996;67:242–244.
16. 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.
17. Leitzes AH, Potter HG, Amaral T, et al. Reliability and accuracy of MRI scanogram in the evaluation of limb length discrepancy. *J Pediatr Orthop*. 2005;25:747–749.
18. Shea KG, Stevens PM, Nelson M, et al. A comparison of manual versus computer-assisted radiographic measurement intraobserver measurement variability for Cobb angles. *Spine*. 1998;23:1786–1790.
19. Bozentka DJ, Beredjik PK, Westawski D, et al. Digital radiographs in the assessment of distal radius fracture parameters. *Clin Orthop*. 2002;397:409–413.

