Skeletal Radiol (2012) 41:187–191
DOI 10.1007/s00256-011-1166-7

SCIENTIFIC ARTICLE

Measurement of leg length discrepancy after total hip arthroplasty. The reliability of a plain radiographic method compared to CT-scanogram

Martin Kjellberg · Bariq Al-Amiry · Erling Englund · Göran O. Sjödén · Arkan S. Sayed-Noor

Received: 21 November 2010 / Revised: 23 February 2011 / Accepted: 27 March 2011 / Published online: 14 April 2011

Abstract

Objective To measure the interobserver reliability and intraobserver reproducibility of post total hip arthroplasty (THA) leg length discrepancy (LLD) measurement on radiographs as well as to evaluate its accuracy by comparing it with LLD measurement on computed tomographic scanogram (CT-scanogram).

Materials and methods In this prospective study, postoperative LLD measurements in ten THA patients were made by four observers on anteroposterior radiographs of the pelvis (inter-teardrop line to the tip of lesser trochanter) and compared to LLD measurements made on CT-scanogram scout views of the lower limb. Two observers repeated the LLD measurements on radiographs 8 weeks after the first measurements. The interobserver reliability of the LLD measurement on plain radiographs was evaluated by comparing the measurements of the four observers and the intraobserver reproducibility by comparing the two repeated measurements made by the two observers.

Results We found excellent interobserver reliability (mean ICC 0.83) and intraobserver reproducibility (ICC 0.90 and 0.88) of the LLD measurements on plain radiographs. There was a moderate to excellent agreement, but with wide variation of measurements among the four observers, when plain radiographic measurement was compared with CT-scanogram (ICC 0.58, 0.60, 0.71, and 0.82).

Conclusion Despite the excellent interobserver reliability and intraobserver reproducibility of LLD measurement on radiographs, clinicians should be aware of its limited accuracy when compared to CT-scanogram.

Keywords Leg length discrepancy · Interobserver reliability · Intraobserver reproducibility · Radiographs · CT-scanogram

M. Kjellberg
Department of Orthopaedics, Sundsvall Hospital,
Sundsvall, Sweden
e-mail: martin.kjellberg@lvn.se

B. Al-Amiry
Department of Radiology, Karolinska University Hospital,
Huddinge,
Stockholm, Sweden
e-mail: bariq1976@hotmail.com

E. Englund
Department of Research and Development,
Västernorrland County,
Sundsvall, Sweden
e-mail: erling.enlund@lvn.se

G. O. Sjödén · A. S. Sayed-Noor (✉)
Department of Orthopaedics, Sundsvall Hospital,
85186 Sundsvall, Sweden
e-mail: arkansam@hotmail.com

G. O. Sjödén
e-mail: goran.sjoden@lvn.se

G. O. Sjödén · A. S. Sayed-Noor
Department of Surgical and Perioperative Sciences,
Umeå University,
901 85 Umeå, Sweden

Introduction

Restoration of leg length discrepancy (LLD) is an important goal when performing a total hip arthroplasty (THA). Failure to achieve this may be associated with general patient dissatisfaction [1], gait disorder [2], greater trochanteric pain [3], nerve palsy [4], and aseptic loosening [5]. Therefore, it is essential to use a reliable and cost-effective method to measure LLD.

Numerous radiological measurement methods of LLD have been described in the literature [6]. These methods have different diagnostic accuracy, cost effectiveness, radiation dose, and time consumption. Radiological LLD measurement has been found to be more accurate than clinical measurement [7–10].

Radiographs are commonly used in the evaluation of LLD before and after THA. The perpendicular distance between a line passing through the lower edge of the teardrop points (inter-teardrop line) or the lower edge of ischial tuberosities (bi-ischial line) to the tip of the lesser trochanter is measured on each side and the difference is the LLD [11]. This measurement is achieved on the AP view of pelvis and proximal femur.

On the other hand, computerized tomography scanogram (CT-scanogram) utilizing a single anteroposterior (AP) scout film over the pelvis and entire lower limbs is another modality to measure LLD [12]. The accuracy and interobserver reliability and intraobserver reproducibility of CT-scanogram have been evaluated by previous studies [6, 13, 14] and were found to be very high. These factors make CT-scanogram a reliable method to measure LLD, and it can be used as a reference method to which other methods can be compared. Unfortunately, CT-scanogram is not readily available in all centres, requires prior scheduling, and is more costly than radiographs, making it inappropriate as a routine method when measuring LLD after THA.

The aims of this prospective study were to measure the interobserver reliability and intraobserver reproducibility of LLD measurement on radiographs (inter-teardrop line to the tip of lesser trochanter) as well as to evaluate its accuracy by comparing it with LLD measurement on CT-scanogram.

Materials and methods

The study was approved by the local ethics committee.

Patients

Ten consecutive patients aged 65 to 88 years (mean 76) with primary osteoarthritis of the hip were included. These patients underwent THA (cemented Lubinus SP II Link, Germany) by one surgeon (A.S.N.) between January and March 2008. Patients with secondary osteoarthritis, or previous spinal, pelvic, or lower limb injuries or fractures were excluded.

LLD measurements

Both radiographs (Siemens, Erlangen, Germany) of the pelvis and CT-scanogram scout (GE Health Care, United Kingdom) view were taken on the third postoperative day.

Before imaging, the patient was supine, and a radiology assistant put the patient's feet together in symmetrical internal rotation in order to standardize the measurements. During imaging, the patient's positioning was monitored by the radiology assistant who was protected from the radiation source by a leaded glass screen.

The obtained images were presented digitally, and the PACS system was used for LLD measurement.

The LLD on radiographs was defined as the difference in perpendicular distance in millimeters between a line passing through the lower edge of the teardrop points to the corresponding tip of the lesser trochanter (Fig. 1). A positive LLD value was obtained when the operated limb was longer than the contralateral side, whereas a negative value indicated the opposite. Measurements were calibrated to a radiopaque standardized metal sphere to assess the degree of magnification. A 1-mm precision scale was used.

The LLD on CT-scanogram was measured as the difference in millimeters between the sum of the femoral and tibial lengths on the operated leg versus the contralateral side. The femoral length at the operated side was measured from the upper edge of the acetabular cup to the lower edge of the intercondylar notch and on the contralateral side from the top of the femoral head to the lower edge of the intercondylar notch. We chose the upper edge of the acetabular cup as we thought it represented the top of the removed femoral head by the THA. On both sides, the tibial length was measured from center of the intercondylar eminence to the center of the tibial plafond (Fig. 2). During CT-scanogram imaging, the positioning of the legs was standardized as with the radiographs. A positive LLD value was indicated when the operated limb was longer than the contralateral side, whereas a negative value indicated the opposite. A 1-mm precision scale was used.

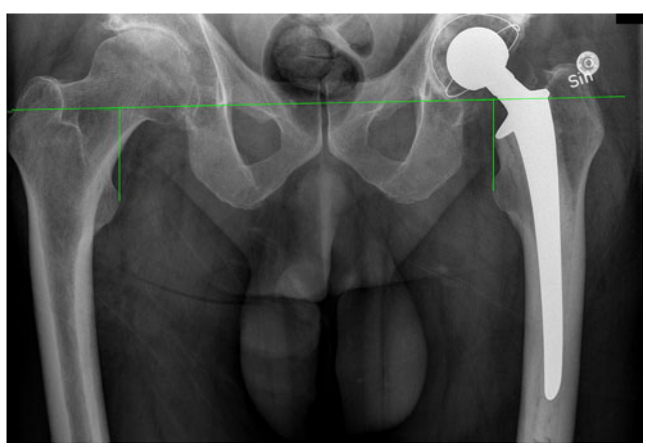

Fig. 1 Radiographic measurement method. The LLD was defined as the difference in perpendicular distance in millimeters between a line passing through the lower edge of the teardrop points to the corresponding tip of the lesser trochanter

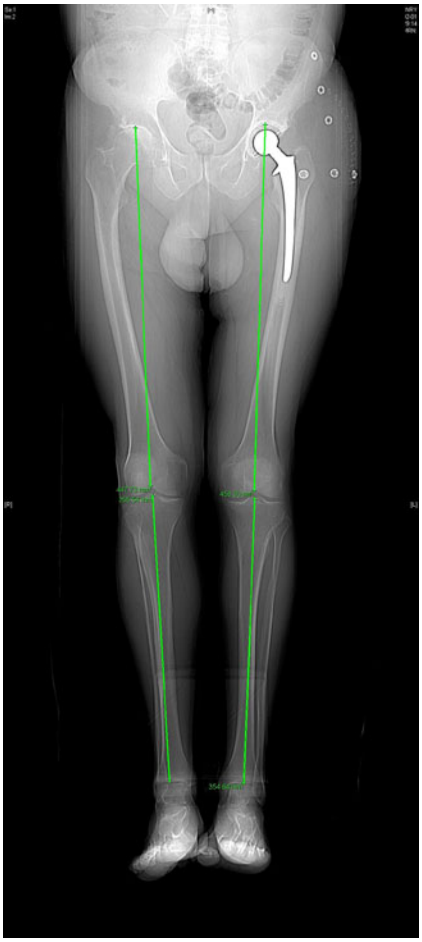

Fig. 2 CT-scanogram measurement method. The LLD was defined as the difference in millimeters between the sum of the femoral and tibial lengths on the operated leg versus the contralateral side. The femoral length at the operated side was measured from the upper edge of the acetabular cup to the lower edge of the intercondylar notch and on the contralateral side from the top of the femoral head to the lower edge of the intercondylar notch. On both sides, the tibial length was measured from the center of the intercondylar eminence to the center of the tibial plafond

Unsatisfactory images showing asymmetry of the hip/lower limb views (evaluated by the bilateral appearance of the obturator foramen, lesser trochanter, and knee and ankle joints) were discarded and replaced with new satisfactory images.

No preoperative plain radiological or CT-scan LLD measurements were evaluated in this study.

The accuracy of the LLD measurements by plain radiographs was evaluated by comparing them to the measurements on CT-scanogram.

LLD measurement analysis

Four observers (one senior orthopaedic resident, one specialist orthopaedic surgeon, one senior radiology resident, and one specialist radiologist) measured LLD on plain radiographs and CT-scanogram. Eight weeks after the first evaluation, one orthopaedic surgeon and one radiologist measured LLD again on radiographs. The observers were blinded to each other's and their own previous results during the course of the study.

The interobserver reliability of the plain radiographic method was evaluated by comparing the measurements of the four observers, while the intraobserver reproducibility was evaluated by comparing the first to the second measurements of the specialist orthopaedic surgeon and the specialist radiologist.

Statistical analysis

SPSS version 16 (SPSS, Chicago, IL, USA) was used for the statistical analysis. The accuracy, interobserver reliability, and intraobserver reproducibility of the tested radiographic method were evaluated by using the intraclass correlation coefficient (ICC) and the Pearson correlation coefficient (r), presented with a 95% confidence interval (CI).

Results

Accuracy of the radiographic method

Table 1 summarizes the ICC with its CI as well as r values of the agreement between the plain radiographic measurements and the CT-scanogram measurements. Observer no. 1 had an excellent agreement (ICC 0.82) between the radiographic and CT-scanogram measurements. Observers no. 2 and 4 had a moderate agreement (ICC 0.58 and 0.60, respectively) while observer no. 3 had a substantial agreement (ICC 0.71). In all cases, there was a wide 95% CI of the obtained measurements.

The interobserver reliability and intraobserver reproducibility of the radiographic method

Table 2 summarizes the interobserver reliability between the radiographic measurements undertaken by the four observers. All observers had excellent interobserver reli-

Table 1 The agreement between the plain radiographic measurements and the CT-scanogram for the four observers

	ICC	95% CI	r
Observer 1	0.82	0.30 to 0.92	0.71
Observer 2	0.58	−0.70 to 0.90	0.50
Observer 3	0.71	−0.18 to 0.90	0.61
Observer 4	0.60	−0.61 to 0.90	0.44

ICC Intraclass correlation coefficient, *CI* confidence interval, *r* Pearson correlation coefficient

Table 2 The interobserver reliability of the radiographic LLD measurement method (four observers) where the mean intraclass correlation coefficient (ICC) was 0.83 (95% confidence interval 0.62–0.95)

	Observer 1	Observer 2	Observer 3	Observer 4
Observer 1	1.00	0.93	0.85	0.79
Observer 2	0.93	1.00	0.89	0.79
Observer 3	0.85	0.89	1.00	0.92
Observer 4	0.79	0.79	0.92	1.00

ability with each other except observer no. 4 who had substantial reliability with observers no. 1 and 2 (ICC 0.79).

For all obtained measurements among the four observers, the mean ICC was 0.83 (CI 0.62–0.95).

Table 3 summarizes the intraobserver reproducibility between the first and second measurements carried out by observer 1 (an orthopaedic surgeon) and observer 4 (a radiologist). Both observers showed excellent intraobserver reproducibility (ICC 0.90 and 0.88), respectively.

Discussion

The evaluation of accuracy of a given measurement method involves the comparison with a gold standard. Furthermore, the ideal measurement method should be reliable and reproducible. No consensus is present in the literature about the gold standard method to measure LLD. Radiographic measurement has been found to be more accurate compared to clinical measurement [7–10]. Leitzes et al. used the electronic calliper measurement as a gold standard in their study [15]. In this study, we compared the plain radiographic measurement method with CT-scanogram owing to its high accuracy and reliability.

For the statistical analysis, the ICC is an appropriate measure for reliability and reproducibility studies involving continuous data (millimeters in the present study), compared with Cohen's kappa, which is used for categorical data. The interpretation of ICC is however controversial. According to Hornij [16], values exceeding 0.75 represent excellent agreement, 0.4–0.75 fair to good agreement, and values less than 0.4 poor agreement. Rheault et al. [17] used the criteria recommended by Landis and Koch [18] to interpret ICC (0.00–0.20 slight agreement, 0.21–0.40 fair agreement, 0.41–0.60 moderate agreement, 0.61–0.80 substantial agreement, and 0.81–1.00 excellent agreement). Others [19] considered the value of 0.60 as a limit of acceptability for application in clinical practice. We chose to use the criteria recommended by Landis and Koch [18].

Table 3 The intraobserver reproducibility of the plain radiographic LLD measurement method for observers 1 and 4

	ICC	95% CI	*r*
Observer 1	0.90	0.63–0.97	0.90
Observer 4	0.88	0.60–0.97	0.88

ICC Intraclass correlation coefficient, *CI* confidence interval, *r* Pearson correlation coefficient

In the present study, we found that the four observers achieved fair to substantial agreement when radiographic measurement was compared with CT-scanogram (Table 1). However, there was a wide 95% confidence interval for each observer. This reflects the limited accuracy of the radiographic LLD measurement method. One possible explanation is that the plain radiographs evaluate LLD at the pelvic level while the CT-scanogram measures LLD for the entire lower limbs. The latter method can therefore be affected by any factor affecting the limb length distal to the lesser trochnater.

The interobserver reliability and intraobserver reproducibility of the radiographic method were excellent. This agrees with Woolson et al. [20] and with White and Dougall [21] who found an excellent interobserver reliability and intraobserver reproducibility for this method, respectively. However these two works [20, 21] were designed to study the effect of LLD on the THA results and not the reliability of the LLD measurement methods used, as in the present study. Moreover, the present study evaluated the accuracy of the radiographic method by comparing it with CT-scanogram. We are unaware of any previous study that evaluated this aspect.

In a newly published study [11], Meermans et al. evaluated LLD measurement preoperatively when templating for THA. The authors found that LLD measurement by using the interteardrop line correlated with LLD measurement by full-leg radiographs better than using the bi-ischial line. The teardrop points have previously been found to be reliable and constant landmarks of the pelvis [22], due to the vertical and rotational stability of these points in association with different pelvis positions.

The present study has some limitations. The sample size consisted of ten patients. Audigé et al. [23] and Slongo et al. [24] mentioned that a sample size of ten subjects is needed when testing the reliability of a new measurement method such as a new classification. We chose to test the interobserver reliability among four observers instead of two to improve the reliability of our results. At the same time, the rotation of the lesser trochanter and the position of the tip can vary according to the rotation of the leg when taking the radiographs. In order to minimize this variation, the legs were kept with a symmetrical internal rotation by the same radiology assistant to make measurements as standardized as possible. The radiographs were accepted for inclusion when obturator foremen, lesser trochanter, and knee and ankle joint all appeared symmetrical bilaterally. Two unsatisfactory CT-scanogram images with asymmetrical rotation were discarded and replaced with new satisfactory images.

Conclusion

The measurement of LLD in THA on radiographs is characterized by excellent interobserver reliability and intraobserver reproducibility. However, the accuracy of this method is limited when compared with CT-scanogram. Clinicians should be aware of this limitation when using the radiographic method in their clinical practice. When LLD evaluation can give rise to further intervention in THA patients such as the need for implant revision, the authors recommend the use of a measurement method with high accuracy and reliability such as CT-scanogram.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Konyves A, Bannister GC. The importance of leg length discrepancy after total hip arthroplasty. J Bone Joint Surg Br. 2005;87(2):155–7.
2. Tanaka R, Shigematsu M, Motooka T, Mawatari M, Hotokebuchi T. Factors influencing the improvement of gait ability after total hip arthroplasty. J Arthroplasty. 2010;25(6):982–5.
3. Sayed-Noor AS, Sjödén GO. Greater trochanteric pain after total hip arthroplasty: the incidence, clinical outcome and associated factors. Hip Int. 2006;16:202–6.
4. Hofmann AA, Skrzynski MC. Leg-length inequality and nerve palsy in total hip arthroplasty: a lawyer awaits! Orthopedics. 2000;23:943–4.
5. Amstutz HC, Ma SM, Jinnah RH, Mai L. Revision of aseptic loose total hip arthroplasties. Clin Orthop Relat Res. 1982;170:21–33.
6. Sabharwal S, Kumar A. Methods for assessing leg length discrepancy. Clin Orthop Relat Res. 2008;466(12):2910–22.
7. Clarke GR. Unequal leg length: an accurate method of detection and some clinical results. Rheumatol Phys Med. 1972;11:385–90.
8. Cleveland RH, Kushner DC, Ogden MC, Herman TE, Kermond W, Correia JA. Determination of leg length discrepancy. A comparison of weight-bearing and supine imaging. Invest Radiol. 1998;23(4):301–4.
9. Lampe HH, Swierstra BA, Diepstraten FM. Measurement of limb length inequality. Comparison of clinical methods with orthoradiography in 190 children. Acta Orthop Scand. 1996;67(3):242–4.
10. Terry MA, Winell JJ, Green DW, Schneider R, Peterson M, Marx RG, et al. Measurement variance in limb length discrepancy: clinical and radiographic assessment of interobserver and intraobserver variability. J Pediatr Orthop. 2005;25(2):197–201.
11. Meermans G, Malik A, Witt J, Haddad F. Preoperative radiographic assessment of limb-length discrepancy in total hip arthroplasty. Clin Orthop Rel Res 2010 [Epub ahead of print]. doi: 10.1007/s11999-010-1588-x
12. Kogutt MS. Computed radiographic imaging: use in low-dose leg length radiography. AJR Am J Roentgenol. 1987;148:1205–6.
13. Huurman WW, Jacobsen FS, Anderson JC, Chu WK. Limb-length discrepancy measured with computerized axial tomographic equipment. J Bone Joint Surg Am. 1987;69:699–705.
14. Poutawera V, Stott NS. The reliability of computed tomography scanograms in the measurement of limb length discrepancy. J Pediatr Orthop B. 2010;19:42–6.
15. Leitzes AH, Potter HG, Amaral T, Marx RG, Lyman S, Widmann RF. Reliability and accuracy of MRI scanogram in the evaluation of limb length discrepancy. J Pediatr Orthop. 2005;25(6):747–9.
16. Horneij E, Holmström E, Hemborg B, Isberg PE, Ekdahl C. Inter-rater reliability and between-days repeatability of eight physical performance tests. Adv Physiother. 2002;4:146–60.
17. Rheault W, Albright B, Byers C, Franta M, Johnson A, Skowronek M, et al. Intertester reliability of the cervical range of motion device. J Orthop Sports Phys Ther. 1992;15:147–50.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
19. Lindell O, Eriksson L, Strender L-E. The reliability of a 10-test package for patients with prolonged back and neck pain: could an examiner without formal medical education be used without loss of quality? A methodological study. BMC Musculoskelet Disord. 2007;8:31.
20. Woolson ST, Hartford JM, Sawyer A. Results of a method of leg-length equalization for patients undergoing primary total hip replacement. J Arthroplasty 1999;14:159–164
21. White TO, Dougall TW. Arthroplasty of the hip. Leg length is not important. J Bone Joint Surg Br. 2002;84:335–8.
22. Goodman SB, Adler SJ, Fyhrie DP, Schurman DJ. The acetabular teardrop and its relevance to acetabular migration. Clin Orthop 1988;236:199–204
23. Audigé L, Bandari M, Kellman J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. Acta Orthop Scand. 2004;75:184–94.
24. Slongo T, Audigé L, Lutz N, Frick S, Schmittenbecher P, Hunter J, et al. Documentation of fracture severity with the AO classification of pediatric long-bone fractures. Acta Orthop. 2007;78:247–53.