Clin Orthop Relat Res (2011) 469:188–199
DOI 10.1007/s11999-010-1447-9

BASIC RESEARCH

Reliability and Agreement of Measures Used in Radiographic Evaluation of the Adult Hip

Nicholas H. Mast MD, Franco Impellizzeri PhD, Stephan Keller MD, Michael Leunig MD

Received: 10 November 2009 / Accepted: 14 June 2010 / Published online: 2 July 2010

Abstract

Background Several mechanical derangements reportedly contribute to the development of noninflammatory arthritis of the hip. Diagnosis of these derangements involves the use of specific radiographic measures (eg, alpha angle, lateral center edge angle, cross-over sign). The reliability of some of these measures is not known, whereas others have not been confirmed.

Questions/purposes We examined the reproducibility of 20 radiographic parameters of the hip used in clinical practice.

Methods Twenty radiographic parameters on standardized digital AP and cross-table lateral radiographs were evaluated by two observers on two different occasions. The parameters were evaluated from the standpoint of reproducibility (reliability and agreement). The intraclass correlation coefficient (ICC), kappa coefficient, and standard error of measurement were calculated. The minimal detectable change was calculated where possible.

Results Interrater reliability ranged from 0.45 to 0.90 for ICC depending on the measure. Intrarater reliability ranged from 0.55 to 0.99. Measurements that could be measured directly (femoral head diameter) were more reliable than measurements requiring estimation on the part of the observer (Tönnis angle, neck-shaft angle). Categorical parameters had interrater and intrarater reliability kappa values greater than 0.90 for all parameters measured. Agreement between repeated measurements, as given by the minimal detectable change, showed many parameters with low absolute reliability have clinical use in the context of the large changes seen in clinical practice.

Conclusion Radiographic hip measures show clinical utility when evaluated from the perspective of agreement and reliability.

Clinical Relevance All measures investigated show clinical utility when evaluated from the perspective of reliability and agreement.

Level of Evidence Level III, diagnostic study. See Guidelines for Authors for a complete description of levels of evidence.

One or more of the authors (NHM) received financial support from AO North America (Paoli, PA, USA).
Each author certifies all investigations related to this study were conducted in conformity with ethical principles of research.
This work was performed at Schulthess Klinik, Zurich, Switzerland.

N. H. Mast
AONA Martin Allgöwer Fellowship, Paoli, PA, USA

F. Impellizzeri, S. Keller
Schulthess Klinik, Zurich, Switzerland

M. Leunig
Department of Orthopedics, Schulthess Clinic, Zurich, Switzerland

N. H. Mast (✉)
c/o SOAR Orthopedics, 500 Arguello, Suite #100, Redwood City, CA 94063, USA
e-mail: doctormast@yahoo.com

Introduction

During the past 10 years a pathomechanical explanation for primary osteoarthritis of the hip has been proposed [10]. The diagnosis of osteoarthritis relies heavily on radiographic imaging of the hip and pelvis [4, 7, 21, 22, 27, 29, 38, 39]. Many radiographic measures have been advanced to help the clinician in this process [7, 11, 14, 15, 21, 24, 27, 39]. Commonly, a selection of measures is used to support or exclude an underlying diagnosis. These same

measures also may be used to measure the adequacy of surgical treatment [4, 32, 34]. Therefore, the process of obtaining and interpreting radiographs in an accurate and reproducible fashion is paramount to the diagnosis, treatment, and study of adult hip disease.

Numerous studies [8, 12, 14, 15, 26, 34, 36, 40] report the interrater (Table 1) and intrarater (Table 2) reliability of adult radiographic hip measures either singularly or collectively to better understand their limitations. One argued radiology of the hip has limited reliability [8], whereas others have argued for good reliability [12, 14, 15, 26, 34, 36, 40]. One previous study focused mainly on one definition of reliability such as the kappa value or ICC to define the limits of reliability for these measures [18]. In doing so, the evaluation of systemic bias and agreement between measures has been neglected [9]. Some of these same studies [8, 34, 36] have attempted to evaluate the clinical utility of a measure by comparing the kappa or ICC value with generic benchmarks (ie, less than 0.75 as having poor reliability). Such a narrow interpretation of reliability may prematurely discard instruments as having limited reliability when, in fact, they may have important clinical information to convey [9]. A more statistically and clinically useful way of evaluating a measure's utility is to present its reproducibility in terms of reliability and agreement [9].

Measurement agreement can be quantified using methods such as the standard error of measurement or the Bland and Altman plot and gives indications regarding "how well a measure produces the same value on repeated measurements" [3]. However, reliability provides indications on "how well individuals in a group can be distinguished from one another by use of a measure" [3], and can be calculated using the ICC. Reliability is important when using a measure for differentiating between individuals or groups, whereas measurement agreement is particularly useful to understand if a measure can be used for evaluating longitudinal changes as often required in clinical settings. Whether a given value for reliability is acceptable depends on the changes we wish to detect and should not be based only on benchmarks.

Therefore, the aim of our study was to examine the reproducibility of various radiographic parameters commonly used in the diagnosis of hip disorders. Specifically, we examined the (1) bias; (2) reliability; and (3) agreement between and within observers; and then (4) used agreement parameters to determine the minimal detectable change (MDC) of the radiographic variables in relation to the corresponding changes considered clinically important.

Materials and Methods

Considering 0.75 as the minimal acceptable value for the ICC and an expected ICC value of 0.90, a sample size of $n > 25$ for a test-retest design ($k = 2$) was estimated ($\alpha = 0.05$ and $\beta = 0.20$) [43]. Using a computerized database of the senior author (ML), preoperative images of 39 patients were randomly selected by alphabetically sorting a consecutive case series of more than 200 cases. The database comprised patients undergoing hip arthroscopy with the diagnoses of cam and/or pincer femoroacetabular impingement (FAI) with various degrees of labral and/or cartilage disorders. Mild dysplasias were observed in this group, although no patient had joint subluxation. The patient population was generally young with an average age of 34.7 years (range, 18–57 years; SD, 12.7). This study was performed as part of a larger investigation to evaluate arthroscopic management of FAI.

All radiographs were taken at the host institution using a standardized protocol. Radiographs included an AP radiograph of the pelvis standardized for rotation and flexion and a cross-table lateral view of the hip in internal rotation. Evaluation of the radiographs was performed on a digital imaging software system (JiveX, Version 4.3.0.1; Visus Technology, Bochum, Germany). Digital measurements were made of each parameter. Scalar values were used whenever possible.

The reviewers included a third-year resident surgeon (SK) and a second-year hip fellow (NM). Both were instructed in how to take the various measurements of the hip and pelvis. Relevant definitions and citations were provided to each observer for reference during the study. The instructions were reviewed again after the surgeons had measured approximately 20 radiographs. A reasonable concordance of measurements was expected after this process. Each reviewer then was presented with a new series of 39 patients from the database to evaluate.

Measurements were taken at two different settings 1 week apart for each observer. Using the two radiographic views for each patient, 10 interval parameters and 10 nominal parameters were evaluated. Definitions for each measurement are provided below, each with a corresponding reference to its description in the literature if available.

For pelvic rotation (center sacral line to symphysis), the overall rotation was evaluated using the AP radiograph of the pelvis. A line was drawn down the center sacral line; a second line was drawn down the middle of the symphysis pubis parallel to the center sacral line. The absolute value of the distance between these lines defines pelvic rotation and was reported in millimeters. The goal of this measure was to calibrate the radiograph to zero rotation. Using digital radiographs, previous authors [36–40] reported interrater and intrarater reliabilities of 0.91 and 0.96, respectively, with acceptable SDs of this measure of 2.1 mm for males and 2.4 mm for females [40].

Pelvic tilt (sacrococcygeal joint to symphysis) was evaluated using the AP radiograph of the pelvis. The

Table 1. Literature review for interrater reliability

Parameters	Clohisy et al. [8]	Nelitz et al. [26]	Jamali et al. [14]	Tannast et al. [36]	Tannast et al. [40]	Steppacher et al. [34]	Kalberer et al. [15]	Gosvig et al. [12]	Current study	
										MDC
Pelvic rotation	K = 0.21				ICC = 0.91				ICC = 0.59	2.5 mm
Pelvic tilt	K = 0.37				ICC = 0.94		ICC = 0.58		ICC = 0.70	13.8 mm
Femoral head diameter							ICC = 0.94		ICC = 0.90	3.5 mm
Length of acetabulum							ICC = 0.91		ICC = 0.81	6.2 mm
Fossa to ilioischial line (acetabular depth)									ICC = 0.58	6.8 mm
Neck-shaft angle (CCD)		ICC = 0.72							ICC = 0.58	12.2°
Tonnis angle (acetabular inclination)	K = 0.64	ICC = 0.82		ICC = 0.61					ICC = 0.45	9.3°
Sharp's angle		ICC = 0.74							ICC = 0.63	5.0°
Lateral CEA		ICC = 0.85		ICC = 0.92					ICC = 0.73	8.3°
Alpha angle (Notzli et al. [27])								ICC = 0.83	ICC = 0.83	16.6°
Extrusion index		ICC = 0.83		ICC = 0.91					ICC = 0.90	5.0°
Tonnis score	K = 0.59					K = 0.74			AdjK = 0.97	
Femoral rotation									AdjK = 1.0	
Ischial spine sign							ICC = 0.91		AdjK = 1.0	
Cross-over sign			K = 0.628	K = 0.60			ICC = 0.65		AdjK = 0.97	
Sphericity									AdjK = 0.95	
Os acetabulum/rim fracture									AdjK = 1.0	
Femoral bump	K = 0.22								AdjK = 0.97	
Herniation pit									AdjK = 0.97	
Linear indentation sign									AdjK = 0.92	

CCD = center-collum-diaphysis angle; CEA = center edge angle; ICC = intraclass correlation coefficient; AdjK = adjusted kappa; MDC = minimal detectable change.

Table 2. Literature review for intrarater reliability

Parameters	Clohisy et al. [8]	Nelitz et al. [26]	Jamali et al. [14]	Tannast et al. [36]	Tannast et al. [40]	Steppacher et al. [34]	Kalberer et al. [15]	Gosvig et al. [12]	Current study	
										MDC
Pelvic rotation	K = 0.57				ICC = 0.96				ICC = 0.73-0.83	1.5–2.3 mm
Pelvic tilt	K = 0.55				ICC = 0.96		ICC = 0.76		ICC = 0.94–0.99	3.1–4.4 mm
Femoral head diameter							ICC = 0.97		ICC = 0.90–0.96	2.4–3.1 mm
Length of acetabulum							ICC = 0.95		ICC = 0.96–0.97	2.7 mm
Fossa to ilioischial line (acetabular depth)									ICC = 0.86–0.95	2.5–3.9 mm
Neck-shaft angle (CCD)		ICC = 0.76							ICC = 0.94–0.95	4.8–15.9°
Tonnis angle (acetabular inclination)	K = 0.73	ICC = 0.86		ICC = 0.74					ICC = 0.88–0.95	2.0–5.5°
Sharp's angle		ICC = 0.70							ICC = 0.55–0.84	3.4–4.3°
Lateral CEA		ICC = 0.88		ICC = 0.97					ICC = 0.86–0.97	3.0–6.0°
Alpha angle (Notzli et al. [27])								ICC = 0.90	ICC = 0.96–0.98	5.8–8.0°
Extrusion index		ICC = 0.73		ICC = 0.94					ICC = 0.80–0.96	3.5–4.4°
Tonnis score	K = 0.60					K = 0.73			AdjK = 0.95–1.00	
Femoral rotation									AdjK = 1.00–1.00	
Ischial spine sign							ICC = 0.92		AdjK = 1.00–1.00	
Cross-over sign			K = 0.698	K = 0.73			ICC = 0.83		AdjK = 0.95–1.00	
Sphericity									AdjK = 0.92–0.95	
Os acetab/rim fracture									AdjK = 1.00–1.00	
Femoral bump	K = 0.30								AdjK = 0.95–1.00	
Herniation pit									AdjK = 0.87–1.00	
Linear indentation sign									AdjK = 0.95–1.00	

CCD = center-collum-diaphysis angle; CEA = center edge angle; ICC = intraclass correlation coefficient; AdjK = adjusted kappa; MDC = minimal detectable change.

distance between the sacrococcygeal joint and the superior aspect of the symphysis pubis was measured and reported in millimeters. Interrater and intrarater reliabilities were reported as 0.94 and 0.96, respectively [40]. Gender-dependent mean values were presented as 32.3 mm for males and 47.3 mm for females [40]. The clinical goal is to standardize radiographs to this measure. Acceptable SDs were reported as 9.2 mm in males and 7.5 mm in females [40].

For acetabular abduction angle (Sharp's angle [30]), using the AP radiograph of the pelvis, one line intersecting both radiographic teardrops was used to define the transverse axis of the pelvis. Another line was drawn from the most inferior portion of the teardrop to the most lateral point of the acetabular rim. The acute angle at the intersection of these two lines measured in degrees defines Sharp's angle. The ICCs for interrater and intrarater reliabilities were reported as 0.74 and 0.70, respectively [26]. A dysplastic acetabulum is commonly defined by an acetabular abduction angle greater than 43° [42].

The fossa to the ilioischial line (acetabular depth), measured on the AP radiograph of the pelvis, is the distance between the deepest portion of the acetabular floor and the ilioischial line measured in millimeters. This is a direct measurement of acetabular depth. Positive values are defined as when the acetabular floor is lateral to the ilioischial line and negative values are defined when the acetabular floor is medial to the ilioischial line.

Extrusion index, measured on the AP radiograph of the pelvis, is defined as the portion of the femoral head uncovered by the acetabular roof divided by the diameter of the femoral head. Femoral head diameter was measured using the Mose template supplied with the digital software. One line was drawn perpendicular to the transverse pelvic axis tangent to the lateral aspect of the femoral head. The other line is drawn in a similar fashion to the lateral edge of the acetabular rim. The distance between these two parallel lines defines the amount of femoral head uncovered. The value is commonly expressed as a percentage; values greater than 25% are typically indicative of dysplasia [21].

Length of the acetabulum, measured on the AP radiograph of the pelvis, is the absolute distance between the most inferior portion of the radiographic teardrop to the lateral acetabular rim measured in millimeters.

Femoral head diameter is measured on the AP radiograph of the pelvis. Using the Mose template [23], the femoral head is matched and the diameter measured in millimeters.

For neck-shaft angle (CCD), measured on the AP radiograph of the pelvis, lines representing the femoral neck and diaphyseal axes are drawn. The neck-shaft angle or CCD angle is the angle subtended by the intersection of these lines. The line of the neck axis is defined by two points: the center of the femoral head as defined using a Mose template [23] and the midpoint of the femoral neck at its isthmus. The line of the shaft axis is defined by a line connecting the midpoints of two lines drawn perpendicularly across the diaphysis of the femur. The angle is reported in degrees. Measurement error has been reported as ± 2° [16]. Coxa vara is defined as a CCD less than 126° and coxa valga greater than 139° [42]; between 126° and 139° is considered clinically normal.

For lateral center-edge angle [44], measured on the AP radiograph of the pelvis, a line is drawn through the center of the femoral head perpendicular to the transverse pelvic axis. Another line is drawn through the center of the femoral head passing through the most superolateral point of the acetabular roof. The angle created by these two lines defines the lateral center-edge angle. It is reported in degrees. ICCs for intrarater and interrater reliabilities are reported as 0.85 and 0.51 [35]. Slightly better ICCs were reported for interrater reliability (0.67) when observers were experienced in evaluating the hip [45]. Dysplasia of the hip is commonly reported as having a center-edge angle less than 20° [24, 25, 32, 42].

Tönnis angle (variously termed acetabular index, acetabular inclination, or horizontal toit externe [41]), measured on the AP radiograph of the pelvis, is defined by the angle created by the line connecting the inferior and lateral aspects of the acetabular sourcil and the transverse pelvic axis. It is reported in degrees with intrarater and interrater reliabilities (ICC) of 0.74 and 0.61, respectively [36]. Clinical range is typically from 4° to 10° [21], whereas values greater than 10° or less than 0° are commonly referred to as abnormal [8].

In the current study, the alpha angle [27] was measured on the cross-table lateral radiograph with internal rotation, although this measure has been described for the frog lateral [22], Dunn lateral [22] and radial MRI cuts [28]. We used the Mose template to best match the femoral head. A line was drawn down the axis of the femoral neck (described previously) with a second line passing through the center of the femoral head to the point where the neck or head leaves the Mose template. A value greater than 42° suggests a head-neck offset deformity [27]. Interrater and intrarater reliabilities, by unpaired two-tailed t test, have been reported as 0.88 and 0.95, respectively [22].

Femoral rotation was determined from the AP radiograph of the pelvis. This measure expresses whether the femoral rotation is acceptable for further measurement [16]. We expressed it as either acceptable or not acceptable. Acceptable femoral rotation has superimposition or intersection of the anterior and posterior borders of the greater trochanter. When the femur is ideally aligned, a hockey-stick shadow typically is seen in the region of the

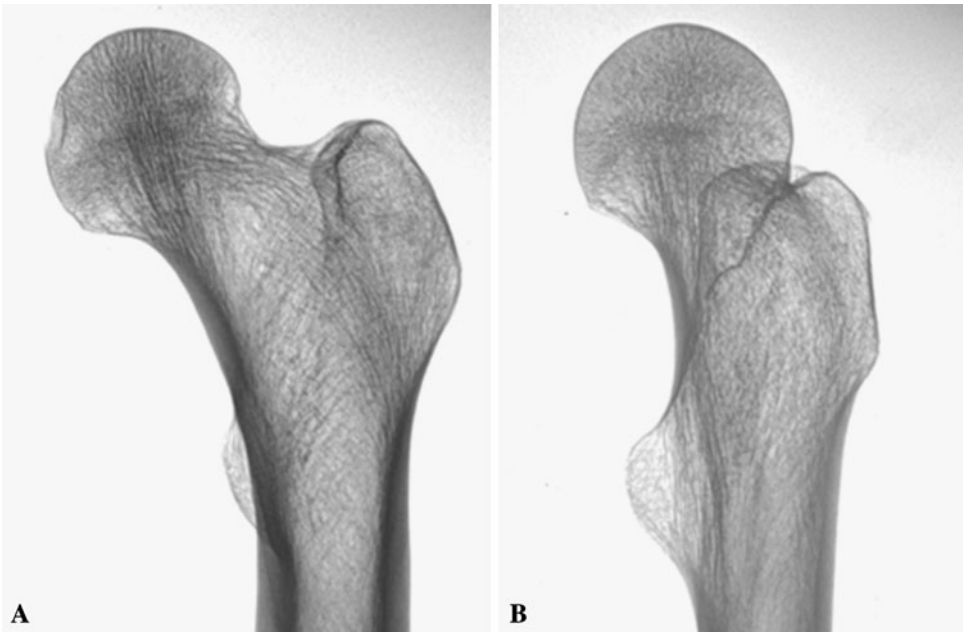

Fig. 1A–B The radiographs show examples of (**A**) acceptable and (**B**) unacceptable femoral rotation. The anterior and posterior borders of the greater trochanter line up in the properly rotated femoral radiograph.

piriformis fossa (Fig. 1). Without these radiographic features, the femoral rotation is considered not acceptable.

The ischial spine sign is a measure of hemipelvic retroversion [15]; this is measured on the AP radiograph of the pelvis standardized for rotation and flexion. The presence of the sign is given by the projection of one or both ischial spines into the pelvic inlet. For the purposes of this study, the sign was either present or absent. This sign, when measured as the length of the spine projecting into the pelvis, has been reported to have interrater and intrarater reliabilities of 0.91 and 0.92, respectively [15].

The cross-over sign [29] is a measure of acetabular retroversion; this is measured on the AP radiograph of the pelvis standardized for rotation and flexion. The presence of the sign is noted as positive when the contour of the anterior rim is located lateral to a corresponding point of the posterior rim contour. It is negative when the contours of the anterior and posterior acetabular rims meet exactly at the most lateral point of the acetabular roof. Kappa values for the cross-over sign have been reported as 0.63 and 0.70 for interrater reliability and 0.67 and 0.70 for intrarater reliability [14].

Femoral head sphericity can be measured on the AP or lateral projections of the hip; both were used in the current study. To assist in our assessment, a Mose template [23] was best fit to the femoral head. If the head deviated from the margin of the reference circle on either view by greater than 2 mm, the femoral head was considered aspherical [7].

Using the AP and lateral radiographs, the Tönnis osteoarthritis grades range from 0 to 3 [41]. Grade 0 indicates there are no radiographic signs of arthritis. Grade 1 includes increased sclerosis of the head and acetabulum, slight joint space narrowing, and slight lipping at the joint margins; Grade 2, small cysts in the head or acetabulum, moderate joint-space narrowing, and moderate loss of sphericity of the head; and Grade 3, large cysts in the head or acetabulum, joint space obliteration or severe joint space narrowing, severe deformity of the femoral head, or evidence of necrosis. Kappa values have been reported for interobserver reliability (0.74) and intraobserver reliability (0.73) [34].

Os acetabulum/rim fractures are evaluated on the AP radiograph. If visible on either view, this measure was positive.

The femoral bump is [1] a qualitative measure based on the gross appearance of the anterior femoral neck on the lateral radiograph. If the anterior and posterior concavities are symmetric, this measure is defined as absent. If, however, there is a protuberance or convexity of the proximal femoral epiphysis, this is defined as present.

The linear indentation sign (femoral trough [39]) is a qualitative measure based on the gross appearance of the anterior femoral neck. If a femoral bump is absent but there is concavity in the anterior femoral neck, this measure is defined as positive. The presence of an anterior femoral neck trough is believed to be associated with pincer impingement [39].

A herniation pit is the presence of a cyst in the anterior femoral head-neck junction seen on either the AP or lateral radiograph diagnostic for a herniation pit [19].

Consistent deviations in measures between observations or observers (bias) were examined by inspecting the 95% confidence intervals of the differences between

measurements. Reliability was calculated using the ICC (two-way mixed, single measure model: ICC3.1) for radiographic data expressed as continuous variables. Intraclass correlations are correlations often used as reliability coefficients. They are ratios between rating variance to total variance. They compare the covariance of the ratings with the total variance and generally are applied to data that can be expressed as continuous variables. The accuracy of the model used was confirmed according to the classification of Shrout and Fleiss [31]. For categorical data, reproducibility was calculated using the kappa coefficient, a method that compares the observed agreement with perfect agreement corrected for chance [33], ie, it indicates the proportion of agreement beyond that expected by chance. Kappa can be influenced by the case distribution (attribute prevalence) and bias. Therefore, the kappa coefficients were adjusted for prevalence and bias [5, 20]. Agreement was determined by calculating the standard error of measurement (SEM) [9]. The MDC, ie, the smallest individual change that can be considered real and not the result of measurement error, was calculated as SEM $\times \sqrt{2} \times 1.96$. Data are presented as means and SDs. The statistical analyses were performed with SPSS 17 (SPSS Inc, Chicago, IL, USA) and an Excel spreadsheet (Excel 2007; Microsoft Inc, Redmond, WA, USA) [20].

Results

We found no evidence of bias between either intraobserver (Table 3) or interobserver (Table 4) measurements. Although the 95% confidence interval was skewed toward

Table 3. Intrarater bias for the radiographic parameters measured by Observers 1 and 2

Parameters	Unit	Observer 1 Mean bias (95%CI)	Observer 2 Mean bias (95%CI)
Pelvic rotation (CSL to symphysis)	mm	0.5 (0.1 to 0.8)	0.6 (0.1 to 1.2)
Pelvic flexion (SC joint to symphysis)	mm	−0.1 (−1.2 to 1.0)	−2.1 (−5.4 to 1.1)
Femoral head diameter	mm	0.7 (−0.1 to 1.4)	−0.2 (−1.0 to 0.6)
Length of acetabulum	mm	1.0 (0.4 to 1.7)	−0.2 (−1.7 to 1.3)
Fossa to ilioischial line	mm	−0.4 (−1.1 to 0.2)	0.0 (−0.4 to 0.4)
Neck-shaft angle	Degrees	−1.1 (−3.6 to 1.6)	1.7 (−0.3 to 3.8)
Tonnis angle	Degrees	0.3 (0.0 to 0.7)	1.4 (−0.1 to 3.0)
Sharp's angle	Degrees	−0.8 (−1.4 to 2.5)	−0.2 (−1.0 to 0.5)
Lateral CEA	Degrees	−0.4 (−0.9 to 0.1)	−0.5 (−1.8 to 0.9)
Alpha angle (lateral)	Degrees	2.4 (1.4 to 3.3)	1.8 (−1.0 to 4.5)
Extrusion index	NA	0.5 (−0.1 to 1.1)	0.7 (−0.1 to 1.4)

CSL = center sacral line; SC = sacrococcygeal; CEA = center edge angle; CI = confidence interval; NA = not applicable.

Table 4. Interrater bias for the radiographic parameters

Parameters	Unit	Observer 1		Observer 2		Difference
		Mean	SD	Mean	SD	
Pelvic rotation (CSL to symphysis)	mm	2.7	2.1	3.3	1.8	0.6 (0.1 to 1.2)
Pelvic flexion (SC joint to symphysis)	mm	38.5	15.4	36.4	9.4	−2.1 (−5.4 to 1.1)
Femoral head diameter	mm	53.6	6.1	53.3	5.1	−0.2 (−1.0 to 0.6)
Length of acetabulum	mm	69.8	6.8	69.6	7.7	−0.2 (−1.7 to 1.3)
Fossa to ilioischial line	mm	−0.2	3.6	1.3	3.9	1.6 (0.4 to 2.7)
Neck-shaft angle	Degrees	129.5	7.4	131.2	6.0	1.7 (−0.3 to 3.8)
Tonnis angle	Degrees	5.3	5.5	6.7	3.2	1.4 (−0.1 to 3.0)
Sharp's angle	Degrees	39.6	2.9	39.9	3.1	0.3 (−0.5 to 1.2)
Lateral CEA	Degrees	31.9	5.7	31.4	5.8	−0.5 (−1.8 to 0.9)
Alpha angle (lateral)	Degrees	56.7	13.8	58.4	15.2	1.8 (−1.0 to 4.5)
Extrusion index	NA	52.7	6.1	52.5	5.1	−0.2 (−1.1 to 0.6)

CSL = center sacral line; SC = sacrococcygeal; CEA = center edge angle; SD = standard deviation; NA = not applicable.

the positive or negative values in various measures, because it straddled zero for most measurements, no significant bias was observed.

ICC values for intrarater reliability showed similar ranges for Observer 1 (Table 5) and Observer 2 (Table 6). These ranged from 0.55 (Sharp's angle) to 0.99 (pelvic flexion). ICC values for interrater reliability ranged from 0.45 (Tönnis angle) to 0.90 (extrusion index, femoral head diameter) (Table 7). Kappa values for interrater reliability showed near perfect agreement for the ischial spine sign, linear indentation sign, and presence or absence of an os acetabulum or rim fracture (Table 8). For the cross-over sign and sphericity, the adjusted kappa coefficients were greater than 0.90. The least reliable parameter examined was the presence of a femoral herniation pit with an adjusted kappa value of 0.87 (Table 8). Without adjustment for the very low prevalence in the sample population, the kappa value was 0.54. Intrarater reliability was greater than interrater reliability for most measures.

Agreement measures such as the SEM and MDC showed that differences between repeated measures were very small. Intraobserver SEM was nearly identical for Observer 1 (Table 5) and Observer 2 (Table 6), with most errors of measurement being between 1 and 2 mm/degrees per observation. Only in the neck-shaft angle did we note a larger SEM for Observer 1. Interobserver SEMs showed higher errors of measurement in alpha angle (6°), neck-shaft angle (4.4°), Tönnis angle (3.4°), and pelvic flexion (7 mm) (Table 7).

The MDC values showed good measurement agreement for most parameters (Tables 3–5). When compared with clinically relevant values (Table 7), the calculated MDC was best for the femoral head diameter using the Mose template included in the software package (3.5 mm for

Table 5. Intrarater reproducibility for the radiographic parameters measured by Observer 1

Parameters	Unit	ICC (95%CI)	SEM	SEM (%)	MDC
Pelvic rotation (CSL to symphysis)	mm	0.83 (0.69 to 0.91)	0.8 (0.6 to 1.0)	25.8 (20.6 to 34.6)	1.5
Pelvic flexion (SC joint to symphysis)	mm	0.94 (0.89 to 0.7)	2.3 (1.9 to 3.0)	5.8 (4.7 to 7.6)	4.4
Femoral head diameter	mm	0.90 (0.81 to 0.95)	1.6 (1.3 to 2.1)	2.9 (2.3 to 3.7)	3.1
Length of acetabulum	mm	0.97 (0.94 to 0.98)	1.4 (1.1 to 1.8)	2.0 (1.6 to 2.6)	2.7
Fossa to ilioischial line	mm	0.86 (0.74 to 0.92)	1.4 (1.2 to 1.8)	29.4 (20.0 to 54.8)	3.9
Neck-shaft angle	Degrees	0.94 (0.89 to 0.97)	5.7 (4.7 to 7.5)		15.9
Tonnis angle	Degrees	0.95 (0.90 to 0.97)	0.7 (0.6 to 1.0)		2.0
Sharp's angle	Degrees	0.84 (0.71 to 0.91)	1.2 (1.0 to 1.6)		3.4
Lateral CEA	Degrees	0.97 (0.93 to 0.98)	1.1 (0.9 to 1.4)		3.0
Alpha angle (lateral)	Degrees	0.98 (0.96 to 0.99)	2.1 (1.7 to 2.7)		5.8
Extrusion index	–	0.96 (0.92 to 0.98)	1.3 (1.0 to 1.6)		3.5

CSL = center sacral line; SC = sacrococcygeal; CEA = center edge angle; ICC = intraclass correlation coefficient; SEM = standard error of measurement; MDC = minimal detectable change.

Table 6. Intrarater reproducibility for the radiographic parameters measured by Observer 2

Parameters	Unit	ICC (95%CI)	SEM	SEM (%)	MDC
Pelvic rotation (CSL to symphysis)	mm	0.73 (0.54 to 0.85)	1.2 (1.0 to 1.5)	94.3 (71.8 to 136.1)	2.3
Pelvic flexion (SC joint to symphysis)	mm	0.99 (0.98 to 1.00)	1.6 (1.3 to 2.0)	5.8 (4.7 to 7.5)	3.1
Femoral head diameter	mm	0.96 (0.92 to 0.98)	1.2 (1.0 to 1.6)	2.3 (1.9 to 3.0)	2.4
Length of acetabulum	mm	0.96 (0.92 to 0.98)	1.4 (1.1 to 1.8)	2.1 (1.7 to 2.7)	2.7
Fossa to ilioischial line	mm	0.94 (0.90 to 0.97)	0.9 (0.8 to 1.2)	29.7 (22.7 to 42.9)	2.5
Neck-shaft angle	Degrees	0.95 (0.90 to 0.97)	1.8 (1.4 to 2.3)		4.8
Tonnis angle	Degrees	0.88 (0.78 to 0.94)	2.0 (1.6 to 2.6)		5.5
Sharp's angle	Degrees	0.55 (0.29 to 0.74)	1.6 (1.3 to 2.0)		4.3
Lateral CEA	Degrees	0.86 (0.77 to 0.93)	2.2 (1.8 to 2.8)		6.0
Alpha angle (lateral)	Degrees	0.96 (0.92 to 0.98)	2.9 (2.4 to 3.7)		8.0
Extrusion index	–	0.80 (0.65 to 0.89)	1.6 (1.3 to 2.1)		4.4

CSL = center sacral line; SC = sacrococcygeal; CEA = center edge angle; ICC = intraclass correlation coefficient; SEM = standard error of measurement; MDC = minimal detectable change.

Table 7. Interater reproducibility for the radiographic parameters

Parameters	Unit	ICC (95%CI)	SEM	SEM (%)	MDC
Pelvic rotation (CSL to symphysis)	mm	0.59 (0.34 to 0.76)	1.3 (1.0 to 1.6)	70.9 (55.0 to 99.5)	2.5
Pelvic flexion (SC joint to symphysis)	mm	0.70 (0.48 to 0.83)	7.0 (5.7 to 9.1)	23.5 (18.8 to 31.4)	13.8
Femoral head diameter	mm	0.90 (0.81 to 0.95)	1.8 (1.5 to 2.3)	2.9 (2.4 to 3.8)	3.5
Length of acetabulum	mm	0.81 (0.66 to 0.90)	3.2 (2.6 to 4.1)	5.8 (4.7 to 17.4)	6.2
Fossa to ilioischial line	mm	0.58 (0.32 to 0.76)	2.4 (2.0 to 3.1)	44.1 (30.0 to 86.0)	6.8
Neck-shaft angle	Degrees	0.58 (0.31 to 0.76)	4.4 (3.6 to 5.7)		12.2
Tonnis angle	Degrees	0.45 (0.15 to 0.67)	3.4 (2.8 to 4.4)		9.3
Sharp's angle	Degrees	0.63 (0.40 to 0.79)	1.8 (1.5 to 2.4)		5.0
Lateral CEA	Degrees	0.73 (0.53 to 085)	3.0 (2.5 to 3.9)		8.3
Alpha angle (lateral)	Degrees	0.83 (0.70 to 0.91)	6.0 (4.9 to 7.7)		16.6
Extrusion index	–	0.90 (0.81 to 0.95)	1.8 (1.5 to 2.3)		5.0

CSL = center sacral line; SC = sacrococcygeal, CEA = center edge angle; ICC = intraclass correlation coefficient; SEM = standard error of measurement; MDC = minimal detectable change.

Table 8. Kappa coefficient for the radiographic parameters measured on nominal scales

Parameters	Categories	Adj k R1	BI	PAI	PA	Adj k R2	BI	PAI	PA	R1 vs R2	BI	PAI	PA
Tonnis score	0,1,2,3*	0.95	0.03	−0.07	97%	1.00	0.00	0.23	100%	0.97	0.01	0.00	98%
Femoral rotation	Yes/No	1.00	0.00	0.64	100%	1.00	0.00	0.49	100%	1.00	0.00	0.56	100%
Ischial spine sign	Yes/No	1.00	0.00	0.43	100%	1.00	0.00	0.69	100%	1.00	0.00	0.56	100%
Cross-over sign	Yes/No	0.95	0.03	0.35	97%	1.00	0.00	0.85	100%	0.97	0.01	0.60	98%
Sphericity	Yes/No	0.95	0.03	−0.71	97%	0.92	0.02	−0.62	97%	0.95	0.03	−0.67	98%
Os acetabulum	Yes/No	1.00	0.00	0.74	100%	1.00	0.00	0.53	100%	1.00	0.00	0.64	100%
Femoral bump	Yes/No	0.95	0.03	-0.05	97%	1.00	0.00	0.23	100%	0.97	0.01	0.09	98%
Femoral herniation pit	Yes/No	0.87	0.08	0.82	92%	1.00	0.00	0.59	100%	0.92	0.04	0.88	96%
Linear indentation sign	Yes/No	1.00	0.00	0.69	100%	0.95	0.03	0.56	97%	0.97	0.01	0.63	97%

BI = bias index; PAI = prevalence asymmetry index; Adj k = adjusted kappa for bias index and prevalance asymmetry index; R1 = Observer 1; R2 = Observer 2; PA = proportion agreement; * the sample only included patients with 0 and 1 classes.

interrater, 2.4–3.1 mm intrarater). The largest MDCs were for the neck-shaft angle (CCD) and alpha angle. For the neck-shaft angle, the interrater MDC was 12.2° and the intrarater MDCs were 4.8° to 15.9°. The alpha angle had a high interrater reliability, whereas the MDC was quite large at 16.6°. Various ICC/kappa values have been published for interrater (Table 1) and intrarater (Table 2) reliabilities. Some of these are in contrast to our findings.

Discussion

The diagnosis of hip pain in an adult is built on a careful history, physical examination, and appropriate imaging. In many cases, the diagnosis hinges on the interpretation of the radiograph. The need exists for reproducible markers for early hip disease. The ideal marker should be reproducible in the context of cross-sectional evaluation of a patient sample (for example, differentiating patients within a study). However, this same marker should be reproducible in a longitudinal evaluation of an individual or group of patients (eg, evaluating the change in a parameter as a response to treatment). The former concept is commonly expressed as the reliability of an index, the latter concept is commonly called the agreement of an index. Although reliability is commonly of interest in the research setting, measurement agreement is paramount in research and clinical settings given the importance of detecting individual and group changes after interventions. Therefore, in the current study, we examined the reproducibility of radiographic parameters commonly used in evaluation of the young adult with hip pain by determining bias, reliability, and agreement. In particular, we interpreted the agreement based on the changes that have clinical relevance.

We note some limitations of this study. First, the entire range of each parameter was not represented. When calculating reproducibility for different variables, it is important to ensure the sample population has an adequate

distribution of all values to make such an analysis meaningful. Our study was of young adults with hip pain and the entire range of values for each parameter was not represented. The Tönnis grade, for example, is a categorical scale used to rate the severity of arthritis. In this population of young adults with hip pain, there were no patients with substantial radiographic arthrosis; all patients had Grade 0 or Grade 1 disease. Similarly, there were parameters measured that had a very low prevalence in our sample and this can influence calculations of the kappa coefficient. However, as mentioned earlier, this was dealt with by calculating adjusted kappa coefficients (Table 8). Although a control group is commonly used when evaluating the reproducibility of diagnoses between practitioners, in the context of evaluating the reproducibility of a given measure, what is more important is that an adequate range of measurements is represented. This is important to better mitigate the effect of prevalence on kappa and ICC. Adding more normal values would only complicate calculations where distributions of abnormal values in the population were small. Second, our sample size is similar to those in some studies [8, 34, 36, 40], but smaller than those in other studies [12, 14, 15, 26]. Despite our small sample size, we believe the current study presents an attempt to evaluate the reproducibility of radiographic data using more comprehensive statistical methods with particular reference to the clinical importance of the findings. Finally, there are concerns regarding whether the findings of reproducibility in our study can be generalized to general orthopaedic surgeons or hip specialists. We concede that radiographic diagnosis relies heavily on the clinician's experience with reading hip radiographs. However, the technical proficiency in actually making the measurements was independent of experience. We observed no significant intraobserver or interobserver bias. If experience was important in the reproducibility of measurements, one would expect evidence of intraobserver or interobserver bias (Table 3). Despite the fact there was no bias observed in our study, thoughtful protocol design would dictate that studies should be designed to minimize the impact of bias. Specifically, one should use the same observer or the same group of observers to evaluate cross-sectional and longitudinal data whenever possible. Similarly, clinical decision-making based on these parameters should take this possible bias into account.

Bias, therefore, not only should be considered during design of studies but also should be measured and reported. The reliability of our measurements is heavily dependent on the presence or absence of bias. The ICC and the kappa coefficient can be heavily influenced by commonly ignored limitations such as heterogeneity of data [13], case distribution (attribute prevalence), number of categories in the measurement scale, nonindependence of rating, and bias [6, 17, 32]. By reporting the prevalence and bias index, an adjusted kappa value can be calculated and reported where appropriate [32].

Using digital imaging software and standardized imaging protocols with control for pelvic tilt and rotation, we found it possible to achieve reasonable intrarater and interrater reliabilities for many of the parameters used in plain radiographic evaluation of the young adult hip. Although the ICC is not directly comparable to the kappa coefficient (except in the specific case of weighted kappa), we found the reliability for many parameters to be better than that reported by Clohisy et al. [8] and more comparable to those of other published reports in the orthopaedic literature [12, 14, 15, 26, 34, 36, 40]. For example, Gosvig et al. [12] reported an interrater ICC of 0.83 and an intrarater ICC of 0.90 for the alpha angle measured on 100 films from the Copenhagen Osteoarthritis Study. Our findings were similar with ICC values of 0.83 and 0.96, respectively. Similarly, Tannast et al. [36] found an interrater ICC of 0.53 (95% confidence interval [CI], 0.34–0.77) for the lateral center-edge angle. This overlapped with our findings of an interrater ICC of 0.73 (95% CI, 0.53–0.85). We had comparable findings to those of three previous studies for the Tönnis angle (AI) with an intrarater ICC range of 0.73 to 0.86 compared with our finding of 0.88 [8, 26, 36]. Our data also showed near perfect reliability for parameters measured on nominal scales with no kappa value less than 0.87. Herniation pits appeared to have a kappa coefficient less than most of the other parameters on nominal scales. These occur in a population of patients with mild deformities and the sometimes subtle radiographic appearance of cysts commonly seen in the superior head-neck junction may have been mistaken for vascular foramina, or vice versa. MRI may be a more reproducible way of determining the presence of herniation pits.

In contrast to reliability, which is used to evaluate a measure used in cross-sectional evaluation of a patient population, agreement is used to evaluate this measure used in a longitudinal fashion. Agreement in this case is defined as the degree to which repeated measurements vary for individuals [2]. To use a concrete example, if one is to measure an angle or distance on a hip radiograph, the amount to which that value varies because of the error of measurement is an expression of the agreement of this measure. Such an analysis of agreement is presented in the context of the SEM [9].

The SEM provides valuable clinical information because it can be used to calculate the MDC, which is defined as the smallest individual change that can be considered real and not the result of measurement error. From the standpoint of agreement, the interrater and intrarater MDCs of most of the parameters were less than the clinically important range (Table 9). In the clinical setting,

Table 9. Clinical ranges as reflected in the literature for various parameters of hip radiographs

Measure	Clinical range for interval parameters			Reference for clinical range
	Normal	Borderline	Pathologic	
Pelvic rotation	0 +/−4.2 mm	N/A	N/A	Tannast et al. [40]
Pelvic tilt				
Male	32.3 mm male +/− 15 mm			
Female	47.3 mm female +/− 15 mm	N/A	N/A	Tannast et al. [40]
Acetabular abduction angle (Sharp)	< 43	N/A	> 43	Tonnis & Heinecke [42]
Fossa to ilioischial line	not deep > 0	N/A	Deep, < 0	Clohisy et al. [7]
Extrusion index	< 25	25	> 25	Mast et al. [21]
Neck-shaft angle (CCD)	126–139	N/A	< 126, > 139	Tonnis & Heinecke [42]
Lateral center edge angle (Wiberg, CEA)	> 25	20–25	< 20	Tonnis & Heinecke [42]
Tonnis angle (acetabular inclination)	0–10	N/A	> 10	Mast et al. [21]
Alpha angle (Notzli et al. [27])				
Male	< 68 males	69–82 males	> 83 males	
Female	< 50 females	51–56 females	> 57 females	Gosvig et al. [12]

CCD = center-collum-diaphysis angle; CEA = center edge angle.

a review of MDC for intraobserver measurements gives some threshold values above which a surgeon's measurements from two radiographs can be considered to show real change. For example, the MDC for one observer measuring the Tönnis angle on two separate occasions ranges from 2.0° (Table 5) to 5.5° (Table 6). As occurs in the situation of a surgeon assessing the adequacy of correction from a postoperative film after periacetabular osteotomy, a change greater than 6° can, in this case, be considered a real change and not the result of measurement error alone. The interrater ICC for this parameter was low (0.45) indicating the Tönnis angle is able only to differentiate patients with very large interindividual differences. Such a low ICC also would imply that for a cross-sectional study, the sample size would have to be approximately five times greater than it would if the ICC were 1 (perfect reliability) [13]. Similar to the interpretation for intraobserver MDC, the MDC for interobserver measurements provides a threshold above which a difference in observations between two surgeons can be considered to represent a real change. This latter concept, however assumes no significant bias between the two surgeons. Differences in training and application of different methods to obtain the same measurements may impact the interobserver bias and thereby influence the interobserver MDC.

For the parameters commonly used in the evaluation of the young adult with hip pain, we found that generally good reliability can be achieved. Using interval scales whenever possible allows for calculation of measurement error in absolute terms (agreement) given by the MDC, which offers a more practical application for these statistical indices beyond those furnished by standard reliability benchmarks. Whether a measure should be used for a clinical or research application should be based on an understanding of the agreement and reliability of a measure of the changes we wish to detect.

Acknowledgments We thank the AO Foundation for their continued support of international collaboration and study. We also thank Anne Mannion PhD, for valuable insight in editing and preparation of the manuscript.

References

1. Anderson LA, Peters CL, Park BB, Stoddard GJ, Erickson JA, Crim JR. Acetabular cartilage delamination in femoroacetabular impingement: risk factors and magnetic resonance imaging diagnosis. *J Bone Joint Surg Am.* 2009;91:305–313.
2. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med* 1998;26:217–238.
3. Bartko JJ. The intraclass correlation coefficient as a measure of reliability. *Psychol Rep.* 1966;19:3–11.
4. Beaulé PE, Allen DJ, Clohisy JC, Schoenecker P, Leunig M. The young adult with hip impingement: deciding on the optimal intervention. *J Bone Joint Surg Am.* 2009;91:210–221.
5. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46:423–429.
6. Chmura Kraemer H, Periyakoil VS, Noda A. Kappa coefficients in medical research. *Stat Med.* 2002; 21:2109–2129.
7. Clohisy JC, Carlisle JC, Beaule PE, Kim YJ, Trousdale RT, Sierra RJ, Leunig M, Schoenecker PL, Millis MB. A systematic approach to the plain radiographic evaluation of the young adult hip. *J Bone Joint Surg Am.* 2008;90(suppl 4):47–66.
8. Clohisy JC, Carlisle JC, Trousdale R, Kim YJ, Beaule PE, Morgan P, Steger-May K, Schoenecker PL, Millis M. Radiographic evaluation of the hip has limited reliability. *Clin Orthop Relat Res.* 2009;467:666–675.
9. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59:1033–1039.

10. Ganz R, Leunig M, Leunig-Ganz K, Harris WH. The etiology of osteoarthritis of the hip: an integrated mechanical concept. *Clin Orthop Relat Res.* 2008;466:264–272.
11. Gosvig KK, Jacobsen S, Palm H, Sonne-Holm S, Magnusson E. A new radiological index for assessing asphericity of the femoral head in cam impingement. *J Bone Joint Surg Br.* 2007;89:1309–1316.
12. Gosvig KK, Jacobsen S, Sonne-Holm S, Gebuhr P. The prevalence of cam-type deformity of the hip joint: a survey of 4151 subjects of the Copenhagen Osteoarthritis Study. *Acta Radiol* 2008;49:436–441.
13. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000;30:1–15.
14. Jamali AA, Mladenov K, Meyer DC, Martinez A, Beck M, Ganz R, Leunig M. Anteroposterior pelvic radiographs to assess acetabular retroversion: high validity of the "cross-over-sign". *J Orthop Res.* 2007;25:758–765.
15. Kalberer F, Sierra RJ, Madan SS, Ganz R, Leunig M. Ischial spine projection into the pelvis: a new sign for acetabular retroversion. *Clin Orthop Relat Res.* 2008;466:677–683.
16. Kay RM, Jaki KA, Skaggs DL. The effect of femoral rotation on the projected femoral neck-shaft angle. *J Pediatr Orthop.* 2000;20:736–739.
17. Kraemer HC. Extension of the kappa coefficient. *Biometrics.* 1980;36:207–216.
18. Landis JR, Koch GG. A one-way components of variance model for categorical data. *Biometrics.* 1977;33:671–679.
19. Leunig M, Beck M, Kalhor M, Kim YJ, Werlen S, Ganz R. Fibrocystic changes at anterosuperior femoral neck: prevalence in hips with femoroacetabular impingement. *Radiology.* 2005;236:237–246.
20. Mackinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement. *Comput Biol Med.* 2000;30:127–134.
21. Mast JW, Brunner RL, Zebrack J. Recognizing acetabular version in the radiographic presentation of hip dysplasia. *Clin Orthop Relat Res.* 2004;418:48–53.
22. Meyer DC, Beck M, Ellis T, Ganz R, Leunig M. Comparison of six radiographic projections to assess femoral head/neck asphericity. *Clin Orthop Relat Res.* 2006;445:181–185.
23. Mose K. Methods of measuring in Legg-Calve-Perthes disease with special regard to the prognosis. *Clin Orthop Relat Res.* 1980;150:103–109.
24. Murphy SB, Ganz R, Muller ME. The prognosis in untreated dysplasia of the hip: a study of radiographic factors that predict the outcome. *J Bone Joint Surg Am.* 1995;77:985–989.
25. Murphy SB, Kijewski PK, Millis MB, Harless A. Acetabular dysplasia in the adolescent and young adult. *Clin Orthop Relat Res.* 1990;261:214–223.
26. Nelitz M, Guenther KP, Gunkel S, Puhl W. Reliability of radiological measurements in the assessment of hip dysplasia in adults. *Br J Radiol.* 1999;72:331–334.
27. Nötzli HP, Wyss TF, Stoecklin CH, Schmid MR, Treiber K, Hodler J. The contour of the femoral head-neck junction as a predictor for the risk of anterior impingement. *J Bone Joint Surg Br.* 2002;84:556–560.
28. Pfirrmann CW, Mengiardi B, Dora C, Kalberer F, Zanetti M, Hodler J. Cam and pincer femoroacetabular impingement: characteristic MR arthrographic findings in 50 patients. *Radiology.* 2006;240:778–785.
29. Reynolds D, Lucas J, Klaue K. Retroversion of the acetabulum: a cause of hip pain. *J Bone Joint Surg Br.* 1999;81:281–288.
30. Sharp IK. Acetabular dysplasia: the acetabular angle. *J Bone Joint Surg Br.* 1961;43:268–272.
31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420–428.
32. Siebenrock KA, Leunig M, Ganz R. Periacetabular osteotomy: the Bernese experience. *Instr Course Lect.* 2001;50:239–245.
33. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005;85:257–268.
34. Steppacher SD, Tannast M, Ganz R, Siebenrock KA. Mean 20-year followup of Bernese periacetabular osteotomy. *Clin Orthop Relat Res.* 2008;466:1633–1644.
35. Tan L, Aktas S, Copuroglu C, Ozcan M, Ture M. Reliability of radiological parameters measured on anteroposterior pelvis radiographs of patients with developmental dysplasia of the hip. *Acta Orthop Belg.* 2001;67:374–379.
36. Tannast M, Mistry S, Steppacher SD, Reichenbach S, Langlotz F, Siebenrock KA, Zheng G. Radiographic analysis of femoroacetabular impingement with Hip2Norm-reliable and validated. *J Orthop Res.* 2008;26:1199–1205.
37. Tannast M, Murphy SB, Langlotz F, Anderson SE, Siebenrock KA. Estimation of pelvic tilt on anteroposterior X-rays: a comparison of six parameters. *Skeletal Radiol.* 2006;35:149–155.
38. Tannast M, Siebenrock KA. Conventional radiographs to assess femoroacetabular impingement. *Instr Course Lect.* 2009;58:203–212.
39. Tannast M, Siebenrock KA, Anderson SE. Femoroacetabular impingement: radiographic diagnosis–what the radiologist should know. *AJR Am J Roentgenol.* 2007;188:1540–1552.
40. Tannast M, Zheng G, Anderegg C, Burckhardt K, Langlotz F, Ganz R, Siebenrock KA. Tilt and rotation correction of acetabular version on pelvic radiographs. *Clin Orthop Relat Res.* 2005;438:182–190.
41. Tönnis D. *Congenital Dysplasia and Dislocation of the Hip in Children and Adults.* New York, NY: Springer; 1987.
42. Tönnis D, Heinecke A. Acetabular and femoral anteversion: relationship with osteoarthritis of the hip. *J Bone Joint Surg Am.* 1999;81:1747–1770.
43. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17:101–110.
44. Wiberg G. Studies on dysplastic acetabula and congenital subluxation of the hip joint: With special reference to the complication of osteoarthritis. *Acta Chir Scand Suppl* 1939;83(suppl 58):1–135.
45. Wiig O, Terjesen T, Svenningsen S. Inter-observer reliability of radiographic classifications and measurements in the assessment of Perthes' disease. *Acta Orthop Scand.* 2002;73:523–530.