Skeletal Radiol
DOI 10.1007/s00256-011-1283-3

SCIENTIFIC ARTICLE

Radiological findings for hip dysplasia at skeletal maturity. Validation of digital and manual measurement techniques

Ingvild Øvstebø Engesæter · Lene Bjerke Laborie · Trude Gundersen Lehmann · Francesco Sera · Jonas Fevang · Douglas Pedersen · José Morcuende · Stein Atle Lie · Lars Birger Engesæter · Karen Rosendahl

Received: 30 June 2011 / Revised: 19 August 2011 / Accepted: 7 September 2011

Abstract
Objective To report on intra-observer, inter-observer, and inter-method reliability and agreement for radiological measurements used in the diagnosis of hip dysplasia at skeletal maturity, as obtained by a manual and a digital measurement technique.
Materials and methods Pelvic radiographs from 95 participants (56 females) in a follow-up hip study of 18- to 19-year-old patients were included. Eleven radiological measurements relevant for hip dysplasia (Sharp's, Wiberg's, and Ogata's angles; acetabular roof angle of Tönnis; articulo-trochanteric distance; acetabular depth-width ratio; femoral head extrusion index; maximum teardrop width; and the joint space width in three different locations) were validated. Three observers measured the radiographs using both a digital measurement program and manually in AgfaWeb1000. Inter-method and inter- and intra-observer agreement were analyzed using the mean differences

I. Ø. Engesæter · L. B. Laborie · T. G. Lehmann · J. Fevang · L. B. Engesæter · K. Rosendahl
Department of Surgical Sciences,
University of Bergen,
Bergen, Norway

L. B. Laborie
e-mail: Lene.Bjerke.Laborie@helse-bergen.no

T. G. Lehmann
e-mail: Trude.Gundersen.Lehmann@helse-bergen.no

J. Fevang
e-mail: Jonas.Meling.Fevang@helse-bergen.no

L. B. Engesæter
e-mail: Lars.Engesaeter@helse-bergen.no

K. Rosendahl
e-mail: Karen.Rosendahl@helse-bergen.no

I. Ø. Engesæter · T. G. Lehmann · J. Fevang · L. B. Engesæter
Department of Orthopaedic Surgery,
Haukeland University Hospital,
Bergen, Norway

I. Ø. Engesæter · L. B. Laborie · K. Rosendahl
Department of Radiology,
Haukeland University Hospital,
Bergen, Norway

F. Sera
Medical Research Council Centre of Epidemiology for Child Health, University College London Institute of Child Health,
London, UK
e-mail: f.sera@ich.ucl.ac.uk

D. Pedersen · J. Morcuende
Department of Orthopaedics and Rehabilitation,
University of Iowa Hospital & Clinics,
Iowa City, IA, USA

D. Pedersen
e-mail: Doug-Pedersen@uiowa.edu

J. Morcuende
e-mail: Jose-Morcuende@uiowa.edu

S. A. Lie
Uni Health, Uni Research,
Bergen, Norway
e-mail: Stein.Lie@uni.no

I. Ø. Engesæter (✉)
The Norwegian Arthroplasty Register, Department of Orthopaedic Surgery, Haukeland University Hospital,
Møllendalsbakken 11, 5021 Bergen, Norway
e-mail: Ingvild.Engeseter@helse-bergen.no

Published online: 25 September 2011

between the readings/readers, establishing the 95% limits of agreement. We also calculated the minimum detectable change and the intra-class correlation coefficient.
Results Large variations among different radiological measurements were demonstrated. However, the variation was not related to the use of either the manual or digital measurement technique. For measurements with greater absolute values (Sharp's angle, femoral head extrusion index, and acetabular depth-width ratio) the inter- and intra-observer and inter-method agreements were better as compared to measurements with lower absolute values (acetabular roof angle, teardrop and joint space width).
Conclusion The inter- and intra-observer variation differs notably across different radiological measurements relevant for hip dysplasia at skeletal maturity, a fact that should be taken into account in clinical practice. The agreement between the manual and digital methods is good.

Keywords Hip dysplasia · Radiological measurements · Validation · Reproducibility

Introduction

Hip dysplasia is a common condition in infancy and childhood and accounts for as many as one-fourth of total hip replacements in patients younger than 40 years of age [1]. The diagnosis is based on the history and physical examination of the patient, supplemented with radiological findings [2, 3]. A correctly performed pelvic radiograph with corresponding measurements is fundamental for the radiological diagnosis. The center-edge (CE) angle of Wiberg [4], Sharp's angle [5], the acetabular depth-width ratio [6, 7] and the femoral head extrusion index [8] are

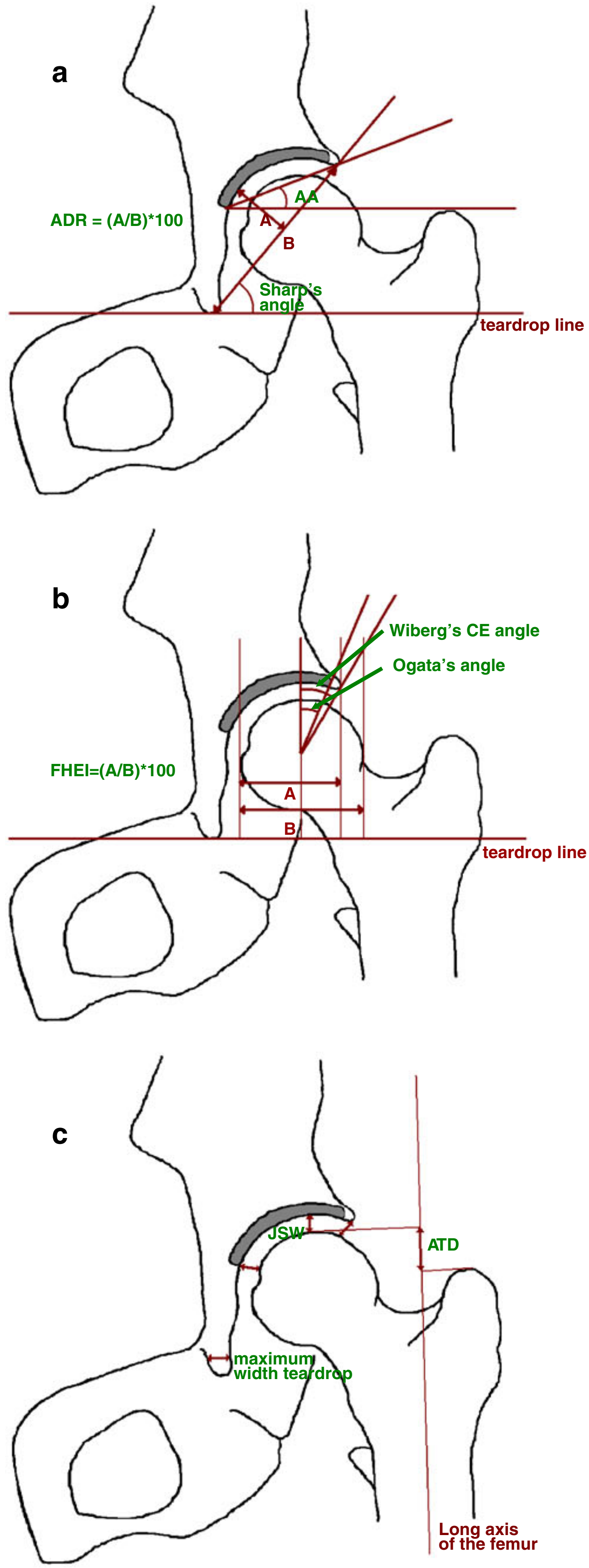

Fig. 1 **a** The acetabular anatomy is assessed by three measurements. (1) Sharp's angle is defined as the angle between the horizontal teardrop line and a line through the inferior teardrop point and the lateral rim of the acetabulum. (2) The acetabular roof angle of Tönnis (AA) describes the angle between the horizontal teardrop line and a line drawn through the medial point of the sourcil and the lateral acetabular rim. (3) The acetabular depth-width ratio (ADR) defines the ratio of the distance between the inferior teardrop point and the lateral acetabular rim, and the depth of the acetabulum. **b** The position of the femoral head to the acetabular cavity was defined using Wiberg's center-edge (CE) angle, a modification of this (Ogata's angle), and the femoral head extrusion index (FHEI). The CE angle was defined as a line through the center of the femoral head and perpendicular to the horizontal teardrop line, and a line running from the center of the femoral head through the lateral acetabular edge (Ogata's angle differs slightly from this by using the lateral edge of the sourcil). FHEI was defined as the percentage of the femoral head lying medial to the lateral acetabular edge. **c** The joint space width (JSW) was measured in three places: medial, middle, and lateral. We also measured the maximum width of the teardrop. The articulotrochanteric distance (ATD) was assessed as the distance between the tangents normal to the long axis of femur to the superior margin of the trochanter major and to the superior contour of the femoral head

commonly used radiographic indices of hip dysplasia. However, the reported accuracy and reproducibility of these radiological markers differ substantially, reflecting differences in imaging techniques and in the definition of measurements used [9–12]. In most PACS systems, several of the measurements used are time-consuming and cumbersome. We therefore set out to investigate the feasibility and repeatability of a digital measurement program for hip dysplasia at skeletal maturity, the "Adult DDH" [12], and to compare the results to manual measurements.

Patients and methods

The radiographs

One hundred pelvic radiographs from a follow-up hip study of 18- to 19-year-old patients were included [mean age 18.3 (SD 0.5) years, 59 females]. All examinations were performed at Haukeland University Hospital (Norway), by one specifically trained radiographer, during the period February 2007 to February 2009 according to the following protocol: erect, anterioposterior (AP) view, feet pointing forward and somewhat parted with the femurs approximately parallel, film-focus distance 1.2 m, centered 2 cm above the pubic symphysis. The dataset was balanced by oversampling dysplastic hips as judged by an experienced pediatric musculoskeletal radiologist (K.R.), who did not take part in the readings. Five out of 100 radiographs were impossible to digitize due to technical problems and were therefore excluded, leaving 95 radiographs for the final analyses.

The pelvic radiographs (DICOM format) were stored within the departmental PACS system before being transferred to a local PC. The images were subsequently opened in the digital measurement program "Adult DDH" (University of Iowa Hospitals and Clinics, Iowa City, Iowa, USA) [12].

All the radiographs (right and left hips) were measured four times (twice using the digital program Adult DDH and twice manually using the AgfaWeb1000 system) by each of two researchers ((L.B.L., I.Ø.E) independently with an interval of at least 3 weeks. Both observers were masked for additional data. A third observer (T.G.L.) measured all images three times, twice using the digital technique and once using the manual AgfaWeb1000 system. Before performing the measurements, all three observers agreed on the precise definitions of landmarks to be used for all the measurements, according to descriptions from the original literature [4–8, 13–19].

Radiographic parameters

Eleven different radiographic measurements were examined. A reference line was drawn between the inferior points of the

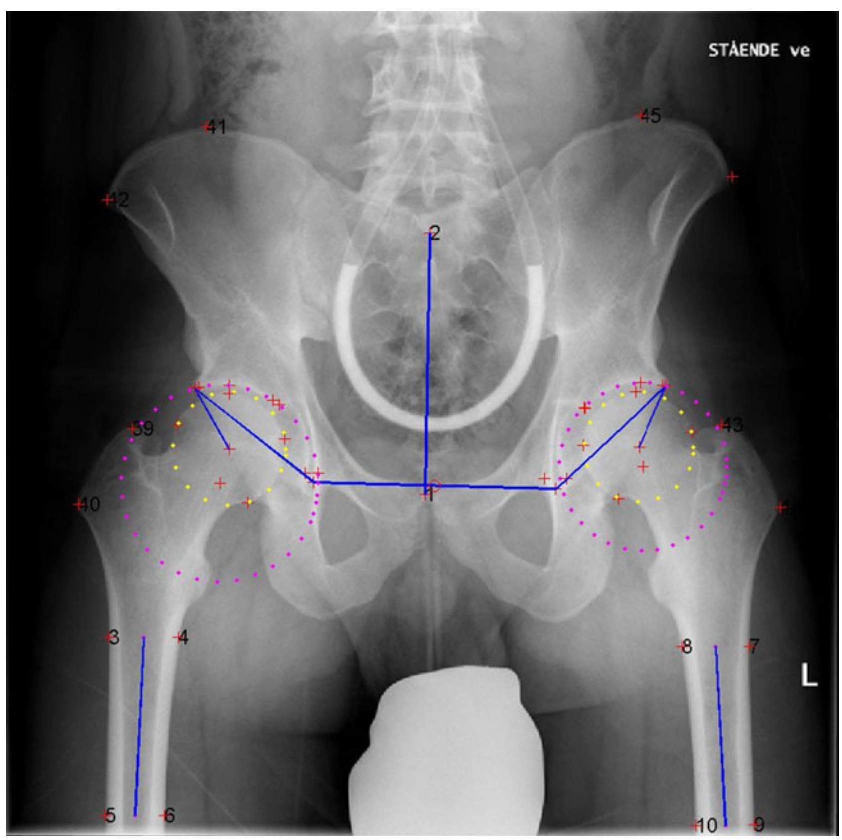

Fig. 2 The radiological landmarks in the digital program Adult DDH were set manually by using a curser. In total 46 landmarks were marked and the results were automatically transferred to an Excel sheet

teardrops (horizontal teardrop line). Sharp's angle [5], the acetabular roof angle of Tönnis (AA) [13–15], and the acetabular depth-width ratio (ADR) [6, 7] describe the acetabular anatomy (Fig. 1a). The position of the femoral head to the acetabular cavity was assessed by the Wiberg's center-edge (CE) angle [4], a modification of this (Ogata's angle) [16], and Heyman and Herndon's femoral head extrusion index (FHEI) [8] (Fig. 1b). We also measured the maximum width of the teardrop [17]; the joint space width medially, laterally, and in the middle [19]; and the articulo-trochanteric distance (ATD) (Fig. 1c). In the digital program, the ATD was assessed as the distance between the tangents normal to the long axis of femur to the superior margin of the trochanter major and to the superior contour of the femoral head [18]. For manual measurements, the ATD was measured as the distance between a horizontal line between the superior margin of the trochanter major on both sides and a perpendicular to this line to the superior margin of the femoral head. In a population with predominately normal hips and where the pelvic radiographs are taken with approximately parallel femurs, there is only a minor difference in these two measuring techniques, and to illustrate this, one of the observers (I.Ø.E.) used both the techniques in the second session of manual measurements.

Digital measurements

Forty-six different radiological landmarks were set manually by using a curser (Fig. 2), allowing the program to estimate the 11 radiological measurements. When measuring short distances such as joint spaces, a standard magnification was applied, allowing for a more precise determination of the measurement points. All measurements were automatically transferred into an Excel sheet at the end of each patient measured.

Manual measurements in AgfaWeb1000

All the manual measurements were performed in AgfaWeb1000 in a standardized order and at a standardized magnification. Both angles and length measurements were estimated using the standard program tools. The center of the femoral head was localized using Mose's templates: a transparent plastic sheet with concentric, high definition circles. The radiological measurements corresponded to those recorded in the digital program, and the measurement values were manually transferred to an Excel sheet during the measuring process.

Statistical analysis

The limits-of-agreement method was used to examine the mean difference between the digital and the manual

Table 1 Intra-observer variability—digital measurements

	Observer 1						Observer 2						Observer 3					
	Mean 1	Mean 2	Mean difference (SD)	95% Limits of agreement	ICC	MDC	Mean 1	Mean 2	Mean difference (SD)	95% Limits of agreement	ICC	MDC	Mean 1	Mean 2	Mean difference (SD)	95% Limits of agreement	ICC	MDC
Sharp's angle	40.74	40.62	0.13 (1.50)	(−2.82; 3.08)	0.92	2.9	40.96	40.70	0.26 (1.45)	(−2.59; 3.10)	0.93	2.9	41.23	41.76	−0.53 (1.52)	(−3.50; 2.45)	0.91	3.1
Wiberg's CE angle	30.07	30.83	−0.76 (2.33)	(−5.32; 3.80)	0.94	4.8	29.46	29.45	0.01 (2.14)	(−4.18; 4.20)	0.95	4.2	28.32	27.75	0.57 (1.85)	(−3.06; 4.21)	0.96	3.8
Ogata's angle	26.53	26.33	0.20 (7.73)	(−14.94; 15.34)	0.93	5.3	28.78	28.86	−0.08 (1.86)	(−3.73; 3.57)	0.97	3.6	25.68	26.65	−0.97 (2.70)	(−6.25; 4.32)	0.93	5.6
Teardrop	7.91	7.70	0.21 (0.87)	(−1.49; 1.91)	0.87	1.7	7.69	7.79	0.10 (1.04)	(−1.93; 2.13)	0.87	2.0	7.05	6.82	0.24 (1.34)	(−2.39; 2.86)	0.78	2.7
AA	6.42	6.32	0.09 (3.15)	(−6.08; 6.27)	0.84	6.1	7.25	6.70	0.54 (2.80)	(−4.95; 6.04)	0.89	5.6	7.21	6.90	0.32 (2.86)	(−5.29; 5.92)	0.90	5.6
ATD	2.19	2.15	0.05 (0.21)	(−0.37; 0.46)	0.93	0.42	2.23	2.29	−0.06 (0.18)	(−0.41; 0.29)	0.94	0.37	2.24	2.22	0.02 (0.14)	(−0.26; 0.30)	0.97	0.28
ADR	295.70	295.01	0.69 (23.44)	(−45.26; 46.64)	0.81	45.7	283.67	282.96	0.71 (15.25)	(−29.17; 30.60)	0.92	29.8	281.09	280.24	0.84 (18.85)	(−0.26; 0.30)	0.89	36.8
FHEI	84.59	84.60	0.00 (2.69)	(−5.28; 5.27)	0.93	5.3	88.16	88.65	−0.49 (4.27)	(−8.87; 7.88)	0.88	8.4	85.17	82.91	2.26 (3.10)	(−3.81; 8.32)	0.88	7.5
JSW medial	4.47	4.64	0.14 (1.18)	(−2.18; 2.45)	0.52	2.3	4.16	4.19	−0.03 (0.75)	(−1.49; 1.43)	0.74	1.5	4.00	3.93	0.07 (1.03)	(−1.94; 2.08)	0.53	2.0
JSW middle	3.36	3.40	−0.04 (0.89)	(−1.79; 1.72)	0.37	1.7	3.80	3.87	−0.08 (0.59)	(−1.24; 1.08)	0.73	1.2	3.89	3.99	−0.10 (0.69)	(−1.46; 1.25)	0.63	1.4
JSW lateral	5.00	4.99	0.01 (1.18)	(−2.31; 2.32)	0.58	2.3	5.95	5.80	0.15 (0.82)	(−1.45; 1.75)	0.80	1.6	5.20	5.49	−0.29 (0.82)	(−1.89; 1.31)	0.75	1.7

AA Acetabular roof angle of Tönnis, *ATD* articulo-trochanteric distance, *ADR* acetabular depth-width ratio, *JSW* joint space width, *ICC* intraclass correlation coefficient, *MDC* minimum detectable change, *SD* standard deviation

measurement techniques (inter-method variability) and also to consider the variability in differences across observers (intra- and inter-observer variability) [20–22]. We obtained the differences between measurements by the two techniques for each individual and calculated the mean and the standard deviations of the difference distribution. When calculating the inter-observer and inter-method differences, we first calculated the mean for each method/observer on each subject and used these pairs of means to compare the two methods/observers. In these cases the standard deviation had to be recalculated as the standard deviation of the differences was too small, because some of the effect of repeated measurement error had been removed. The corrected standard deviation of differences was given by $\sqrt{S_D^2 + \left(1 - \frac{1}{m_1}\right)S_1^2 + \left(1 - \frac{1}{m_2}\right)S_2^2}$, where S_D represented the standard deviation of the mean difference between the methods, S_1 and S_2 the within-subject standard deviation for each method separately, and m_1 and m_2 the number of observations on each subject for the respective methods. The 95% limits of agreement were estimated as mean difference between the two measurements ±1.96 standard deviations (SD). The assumption that the agreement was similar over the range of measurements was checked by plotting the differences against the average of the two methods. The results are given for the right hip.

Replicate measurements by each technique on each individual were obtained to compare repeatability of the two methods [21]. For each technique (manual and digital) and for each observer, the inter-observer reliability was expressed in terms of the intra-class correlation coefficient (ICC) calculated using a one-way random effect ANOVA table [formula ICC (1)] [23]. For each observer, the inter-method reliability was also expressed in terms of ICC calculated using two-way random effect ANOVA table [formula ICC (A,1)] [23]. The ANOVA tables used for ICC indices were used to calculate minimum detectable change (MDC) as indicated by de Vet and colleagues [24]. More in detail $MDC = 1.96 \cdot \sqrt{2} \cdot SEM$ where SEM is the standard error of measurement, calculated as the square root of mean square error (MSE) within ANOVA tables.

The statistical package PASW Statistics 18 for Microsoft Windows® (Microsoft, Redmond, WA) and Stata Statistical Software, Release 11 (StataCorp., 2009, College Station, TX) were used for the statistical analysis.

Ethics approval

The study was conducted in accordance with the ethical standards of the Regional Ethics Committee for Medical and Health Research. The research protocol was approved by the regional ethics committee according to standard regulations. All participants gave written informed consent according to the Helsinki Declaration.

Results

A total of 95 pelvic radiographs were measured manually (five repetitions) and digitally (six repetitions) by three observers, independently. The results of the radiographic measurements with mean, mean difference, standard deviation of the difference, 95% limits of agreement, ICC, and MDC are listed in Tables 1, 2, 3, 4, 5.

For both techniques under investigation, the overall intra- and inter-observer variability differed substantially between the different measurements obtained (Fig. 3) with low levels noted for Sharp's angle (Fig. 4a) and high levels

Table 2 Intra-observer variability—manual measurements

	Observer 1						Observer 2					
	Mean 1	Mean 2	Mean difference (SD)	95% Limits of agreement	ICC	MDC	Mean 1	Mean 2	Mean difference (SD)	95% Limits of agreement	ICC	MDC
Sharp's angle	41.10	40.84	0.25 (2.43)	(−4.51; 5.01)	0.83	4.8	41.22	40.94	0.28 (1.73)	(−3.11; 3.68)	0.89	3.4
Wiberg's CE angle	29.23	29.80	−0.58 (1.46)	(−3.44; 2.28)	0.97	3.1	29.78	30.35	−0.57 (4.12)	(−8.65; 7.50)	0.84	8.1
Ogata's angle	27.04	26.34	0.70 (1.93)	(−3.08; 4.48)	0.96	4.0	28.94	29.80	−0.86 (4.10)	(−8.90; 7.17)	0.85	8.2
Teardrop	7.58	7.55	0.04 (0.43)	(−0.80; 0.88)	0.97	0.8	7.41	7.23	0.18 (1.41)	(−2.59; 2.95)	0.77	2.8
AA	9.02	8.04	0.98 (1.22)	(−1.42; 3.38)	0.95	3.1	6.81	7.28	−0.47 (3.82)	(−7.95; 7.01)	0.80	7.5
ATD	2.32	2.24	0.08 (0.15)	(−0.22; 0.38)	0.95	0.33	2.34	2.34	0.00 (0.09)	(−0.18; 0.18)	0.99	0.18
ADR	296.56	308.42	−11.85 (21.07)	(−53.15; 29.44)	0.72	47.2	290.21	286.14	4.07 (22.57)	(−40.16; 48.30)	0.80	44.7
FHEI	84.04	84.18	−0.14 (2.11)	(−4.28; 4.00)	0.95	4.1	83.58	83.79	−0.21 (4.91)	(−9.83; 9.42)	0.79	9.6
JSW medial	4.70	4.52	0.18 (0.62)	(−1.04; 1.40)	0.79	1.3	4.36	4.22	0.14 (0.71)	(−1.26; 1.54)	0.71	1.4
JSW middle	4.07	3.84	0.23 (0.42)	(−0.58; 1.04)	0.81	0.9	4.09	4.20	−0.11 (0.65)	(−1.39; 1.17)	0.57	1.3
JSW lateral	5.51	5.12	0.39 (0.68)	(−0.93; 1.71)	0.74	1.5	5.47	5.67	−0.20 (0.93)	(−2.03; 1.62)	0.69	1.9

AA Acetabular roof angle of Tönnis, *ATD* articulo-trochanteric distance, *ADR* acetabular depth-width ratio, *JSW* joint space width, *ICC* intraclass correlation coefficient, *MDC* minimum detectable change, *SD* standard deviation

Table 3 Inter-observer variability—digital measurements

	Mean			Mean difference (SD)			95% Limits of agreement			ICC	MDC
	Observer 1	Observer 2	Observer 3	Observer 1-2	Observer 1-3	Observer 2-3	Observer 1-2	Observer 1-3	Observer 2-3		
Sharp's angle	40.68	40.83	41.49	−0.15 (1.77)	−0.81 (2.10)	−0.66 (1.73)	(−3.61; 3.32)	(−4.93; 3.31)	(−4.05; 2.72)	0.89	3.5
Wiberg's CE angle	30.45	29.46	28.04	0.99 (3.27)	2.41 (3.45)	1.42 (2.54)	(−5.41; 7.39)	(−4.36; 9.18)	(−3.55; 6.39)	0.91	5.7
Ogata's angle	26.43	28.82	26.16	−2.39 (2.90)	0.27 (3.02)	2.66 (2.74)	(−8.07; 3.28)	(−5.66; 6.19)	(−2.71; 8.03)	0.93	5.5
Teardrop	7.80	7.74	6.93	0.06 (1.22)	0.87 (1.17)	0.81 (1.26)	(−2.32; 2.44)	(−1.42; 3.16)	(−1.65; 3.27)	0.81	2.3
AA	6.37	6.97	7.05	−0.61 (3.20)	−0.69 (3.40)	−0.08 (2.98)	(−6.87; 5.66)	(−7.35; 5.98)	(−5.92; 5.76)	0.86	6.2
ATD	2.17	2.26	2.23	−0.09 (0.21)	−0.06 (0.19)	0.03 (0.17)	(−0.51; 0.32)	(−0.43; 0.31)	(−0.31; 0.38)	0.95	0.37
ADR	295.36	283.31	280.67	12.04 (23.90)	14.69 (23.74)	2.65 (19.32)	(−34.79; 58.88)	(−31.84; 61.22)	(−35.22; 40.52)	0.84	42.9
FHEI	84.59	88.40	84.04	−3.81 (5.58)	0.56 (4.21)	4.37 (4.70)	(−14.76; 7.14)	(−7.70; 8.81)	(−4.84; 13.58)	0.84	8.9
JSW medial	4.41	4.18	3.97	0.23 (1.07)	0.44 (1.16)	0.21 (1.01)	(−1.86; 2.32)	(−1.84; 2.71)	(−1.78; 2.19)	0.53	2.1
JSW middle	3.38	3.84	3.94	−0.46 (0.89)	−0.56 (0.89)	−0.10 (0.71)	(−2.19; 1.28)	(−2.30; 1.18)	(−1.50; 1.29)	0.49	1.6
JSW lateral	5.00	5.88	5.35	−0.88 (1.05)	−0.35 (1.03)	0.53 (0.92)	(−2.95; 1.19)	(−2.37; 1.67)	(−1.28; 2.34)	0.69	1.9

AA Acetabular roof angle of Tönnis, *ATD* articulo-trochanteric distance, *ADR* acetabular depth-width ratio, *JSW* joint space width, *ICC* intraclass correlation coefficient, *MDC* minimum detectable change, *SD* standard deviation

noted for the AA (Fig. 4b). Measuring time was shorter for the digital (approximately 2 min) compared to the manual technique (approximately 5 min 30 s).

Intra-observer variability

Low levels of intra-observer variability were noted for Sharp's angle and for FHEI, with differences below 10% of their mean values (Tables 1 and 2 and Fig. 4). In general, high levels of variability were noted for smaller values such as the AA, JSW, and maximum teardrop width. The two techniques showed similar patterns of variability across measurements.

Inter-observer variability

The inter-observer variability did not differ substantially from that found within observers and also displayed similar patterns across measurements (Tables 3 and 4). Low levels of inter-observer variability were found for ATD, Sharp's angle, and FHEI, while high levels were seen for AA.

Inter-method variability—digital vs. manual measurements

The inter-method variability (Table 5) did not differ substantially from that found between or within observers. The 95% limits of agreement were narrower for measure-

Table 4 Inter-observer variability—manual measurements

	Mean			Mean difference (SD)			95% Limits of agreement			ICC	MDC
	Observer 1	Observer 2	Observer 3	Observer 1-2	Observer 1-3	Observer 2-3	Observer 1-2	Observer 1-3	Observer 2-3		
Sharp's angle	40.97	41.08	40.61	−0.11 (2.42)	0.36 (2.42)	0.47 (1.92)	(−4.85; 4.63)	(−4.38; 5.10)	(−3.30; 4.23)	0.83	4.4
Wiberg's CE angle	29.51	30.06	30.43	−0.55 (4.29)	−0.91 (4.07)	−0.36 (3.98)	(−8.96; 7.86)	(−8.89; 7.07)	(−8.17; 7.44)	0.84	7.8
Ogata's angle	26.69	29.37	30.06	−2.68 (3.82)	−3.37 (3.48)	−0.69 (4.02)	(−10.17; 4.82)	(−10.19; 3.46)	(−8.56; 7.18)	0.88	7.2
Teardrop	7.57	7.32	6.48	0.25 (1.38)	1.14 (1.00)	0.94 (1.29)	(−2.46; 2.96)	(−0.82; 3.10)	(−1.58; 3.47)	0.79	2.4
AA	8.53	7.05	7.49	1.48 (3.44)	1.04 (2.56)	−0.44 (3.74)	(−5.26; 8.23)	(−3.97; 6.05)	(−7.76; 6.88)	0.83	6.3
ATD	2.28	2.34	2.34	−0.06 (0.14)	−0.07 (0.14)	0.00 (0.08)	(−0.34; 0.21)	(−0.34; 0.20)	(−0.17; 0.16)	0.98	0.24
ADR	302.49	288.17	282.30	14.32 (25.02)	20.19 (24.82)	5.88 (21.26)	(−34.71; 63.35)	(−28.46; 68.85)	(−35.80; 47.55)	0.77	45.4
FHEI	84.11	83.68	81.99	0.43 (4.53)	2.12 (3.89)	1.69 (4.53)	(−8.45; 9.31)	(−5.51; 9.74)	(−7.18; 10.56)	0.82	8.4
JSW medial	4.61	4.29	4.37	0.32 (0.96)	0.24 (0.95)	−0.08 (0.82)	(−1.55; 2.20)	(−1.62; 2.10)	(−1.68; 1.52)	0.57	1.7
JSW middle	3.95	4.15	4.48	−0.20 (0.79)	−0.53 (0.69)	−0.34 (0.70)	(−1.74; 1.35)	(−1.88; 0.81)	(−1.70; 1.03)	0.55	1.4
JSW lateral	5.31	5.57	5.59	−0.26 (0.97)	−0.28 (0.94)	−0.02 (1.00)	(−2.15; 1.64)	(−2.12; 1.56)	(−1.99; 1.95)	0.65	1.8

AA Acetabular roof angle of Tönnis, *ATD* articulo-trochanteric distance, *ADR* acetabular depth-width ratio, *JSW* joint space width, *ICC* intraclass correlation coefficient, *MDC* minimum detectable change, *SD* standard deviation

Table 5 Inter-method variability—digital vs. manual measurements

	Observer 1						Observer 2						Observer 3					
	Mean digital	Mean manual	Mean difference (SD)	95% Limits of agreement	ICC	MDC	Mean digital	Mean manual	Mean difference (SD)	95% Limits of agreement	ICC	MDC	Mean digital	Mean manual	Mean difference (SD)	95% Limits of agreement	ICC	MDC
Sharp's angle	40.68	40.97	−0.29 (2.16)	(−4.53; 3.95)	0.86	4.1	40.83	41.08	−0.25 (1.90)	(−3.98; 3.48)	0.89	3.5	41.49	40.61	0.88 (1.57)	(−2.20; 3.97)	0.89	3.0
Wiberg's CE angle	30.45	29.51	0.94 (2.92)	(−4.80; 6.67)	0.92	5.1	29.46	30.06	−0.61 (3.42)	(−7.31; 6.10)	0.89	6.5	28.04	30.43	−2.39 (2.79)	(−7.86; 3.08)	0.92	4.9
Ogata's angle	26.43	26.69	−0.26 (3.22)	(−6.57; 6.04)	0.92	5.8	28.82	29.37	−0.55 (3.62)	(−7.65; 6.56)	0.89	6.7	26.16	30.06	−3.89 (3.25)	(−10.26; 2.47)	0.89	6.0
Teardrop	7.80	7.57	0.24 (0.95)	(−1.62; 2.10)	0.88	1.7	7.74	7.32	0.42 (1.36)	(−2.24; 3.08)	0.81	2.5	7.00	6.48	0.51 (1.30)	(−2.04; 3.06)	0.73	2.6
AA	6.37	8.53	−2.16 (2.67)	(−7.39; 3.07)	0.88	5.0	6.97	7.05	−0.07 (3.55)	(−7.02; 6.88)	0.85	6.6	7.05	7.49	−0.43 (3.24)	(−6.78; 5.91)	0.83	6.1
ATD	2.17	2.28	−0.11 (0.29)	(−0.68; 0.46)	0.89	0.50	2.26	2.34	−0.08 (0.35)	(−0.77; 0.62)	0.85	0.60	2.23	2.35	−0.12 (0.34)	(−0.78; 0.55)	0.83	0.57
ADR	295.36	302.49	−7.13 (28.43)	(−62.87; 48.60)	0.72	51.2	283.31	288.17	−4.86 (21.82)	(−47.62; 37.91)	0.86	39.1	280.67	282.30	−1.63 (21.36)	(−43.49; 40.23)	0.82	40.3
FHEI	84.59	84.11	0.48 (3.35)	(−6.08; 7.05)	0.91	6.0	88.40	83.68	4.72 (5.29)	(−5.64; 15.09)	0.78	10.6	84.04	81.99	2.05 (4.09)	(−5.98; 10.07)	0.84	7.2
JSW medial	4.41	4.61	−0.21 (1.16)	(−2.47; 2.06)	0.51	2.1	4.18	4.29	−0.11 (1.05)	(−2.18; 1.95)	0.52	1.9	3.97	4.37	−0.40 (1.03)	(−2.43; 1.63)	0.44	2.0
JSW middle	3.38	3.95	−0.57 (0.88)	(−2.30; 1.16)	0.45	1.6	3.84	4.15	−0.31 (0.79)	(−1.86; 1.24)	0.57	1.4	3.94	4.48	−0.55 (0.78)	(−2.08; 0.99)	0.54	1.5
JSW lateral	5.00	5.31	−0.32 (1.12)	(−2.51; 1.88)	0.60	2.1	5.88	5.57	0.31 (1.06)	(−1.78; 2.39)	0.70	1.8	5.88	5.57	0.31 (0.00)	(0.31; 0.31)	0.65	1.9

AA Acetabular roof angle of Tönnis, *ATD* articulo-trochanteric distance, *ADR* acetabular depth-width ratio, *JSW* joint space width, *ICC* intraclass correlation coefficient, *MDC* minimum detectable change, *SD* standard deviation

ments with greater absolute values such as Sharp's angle, FHEI, and ADR.

Discussion

Radiographic measurements are important in diagnosing hip dysplasia. We have shown that the accuracy of commonly used measurements at skeletal maturity is relatively similar for a novel digital technique as compared to the routine, manual technique. However, the measurement variation differed notably across the different radiological measurements. We found Sharp's angle and FHEI to be the more accurate, but acceptable results were also seen for Wiberg's CE angle, ATD, and ADR, with only minor differences according to the technique used.

The strengths of our study are the high numbers of repeated measurements and the number of observers, the thorough standardization process performed prior to study start, the relatively balanced data set, and the different statistical approaches used. All images were measured independently by three observers, a total of 11 times, providing a strong basis for reproducibility analyses. The pelvic radiographs were collected from a population based cohort of 18- to 19-year-olds who were attending a follow-up hip study on hip dysplasia. All examinations were performed within the same radiology department by one specially trained radiographer and according to a carefully discussed protocol. Prior to the study start, the original literature on all measurements was studied in detail, and all measurements were discussed and agreed on. Furthermore, we established a standard procedure of all manual measurements, including standard magnification for small measurement values. By following this detailed standardization process, we were able to reduce potential biases produced by poor image quality and disagreement between the observers on radiological landmarks.

We acknowledge some limitations to our study. First, the research fellows were relatively inexperienced before the study start. However, all had a special interest in hip dysplasia, and the standardization process was supervised and discussed with a senior pediatric radiologist (K.R.) and a senior pediatric orthopedic consultant (L.B.E.). Second, the data set used in the present study was not perfectly balanced in that two-thirds of the cases were subjectively judged as normal, one-third had a mild degree of dysplasia, and only nine had significant dysplasia. None of the hips were subluxated or dislocated.

Several statistical strategies have been described to be used in the evaluation of reproducibility in measurement studies [21, 24, 25], but there is disagreement as to which method is the most appropriate. The term reproducibility

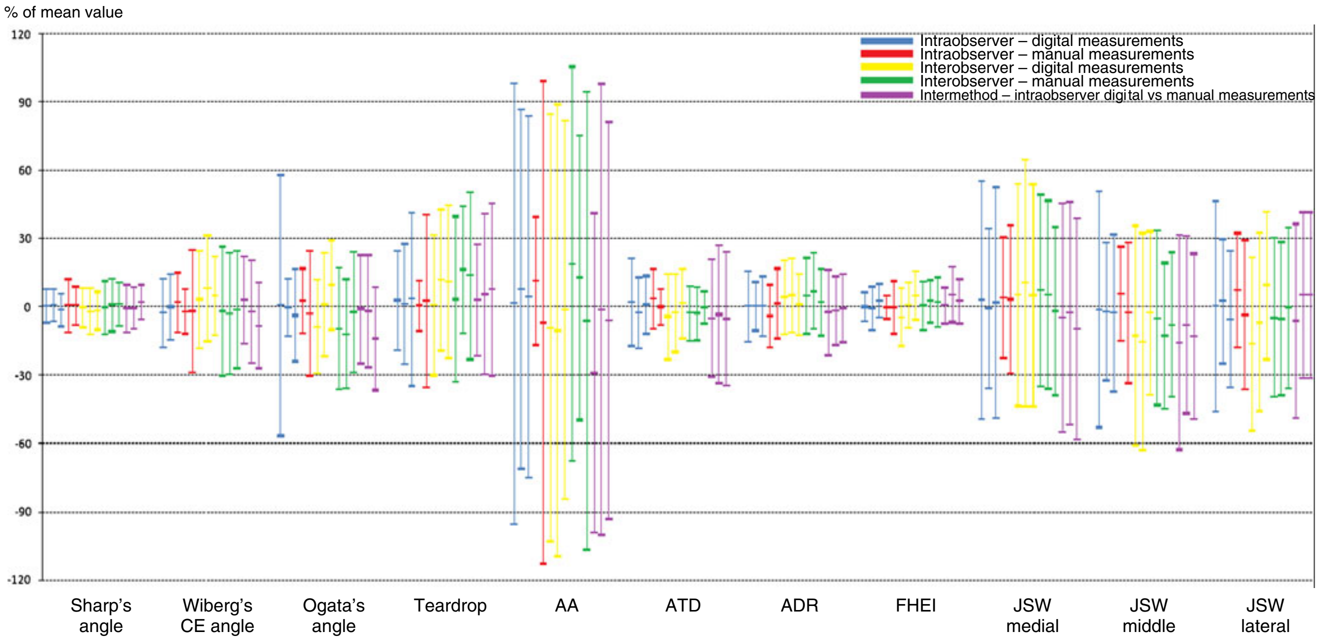

Fig. 3 Overview of measurement variations for the different radiological markers. Each *line* represents the mean difference in percentage of the mean value (mean difference/mean × 100%) with the corresponding 95% limits of agreement in percent [(mean difference/mean × 100%) ± (1.96 SD/mean × 100%)]

includes both agreement and reliability, two terms often used interchangeably in the literature. However, it is important to emphasize that these terms focus on different aspects of reproducibility. The agreement parameters (e.g., the Bland-Altman method and minimum detectable change) are more related to the measurement instrument itself and assess closeness of repeated measurements scores. The reliability parameter (e.g., intra-class correlation coefficient) is related to how well different measurements/patients can be distinguished from each other. A good correlation will for instance be seen for any two methods designed to measure the same parameter and does not automatically imply that there is a good agreement [26, 27]. In general, the Bland-Altman approach, analyzing the difference between measurements by two methods/observers on each subject is the preferred method for analyzing agreement on continuous data [20–22]. To provide estimates of both components (agreement and reliability) of reproducibility, and also to be able to compare our results with others, we chose to include both the Bland-Altman measures of agreement and ICC in the present study. We also included the minimum detectable change (MDC), another parameter of agreement, which is based on the standard error of the measurement (SEM) [24] and gives a measure of detectable clinically relevant change.

The digital measurement program used in the present study was validated by the developers in a previous study including 20 radiographs from patients with established hip dysplasia [12]. The study showed acceptable intra- and inter-observer reliability, as well as inter-method reliability. However, the current study includes almost five times as many radiographs, and the measurements were performed by three independent observers. In our study, the manual measurements were performed in the AgfaWeb1000 as opposed to measurements by hand in the former study.

Radiographic measurements have been subject to validation in numerous studies. For measurements assessing the acetabular anatomy (Fig. 1a), our results compare favorably with those reported by others. In 1999, Nelitz and colleagues measured 100 radiographs from patients with unilateral DDH, aged 16–32 years [9]. They reported acceptable inter-observer ICC values for Sharp's angle (ICC 0.74–0.78) and AA (ICC 0.82–0.88), but inferior values for ADR (ICC 0.58–0.63). Our results agreed well for Sharp's angle (0.83–0.89) and for AA (ICC 0.83–0.86), but we report better results for ADR (ICC 0.77–0.84). Although the inter-observer ICC value for AA was acceptable, the 95% limits of agreement were far too wide for this measurement to be valid, reflecting one of the major problems associated with relying on the ICC alone. Our results show high standard deviations for both manual and digital measurements reflecting the low level of agreement for this specific measurement (Figs. 3 and 4b). Troelsen and colleagues reported similar findings for AA in a study of 25 radiographs from patients referred to the outpatient clinic due to hip pain, aged 15–55 years [11]. For inter-observer measurements, the SD ranged between 1.9 and 4.1°

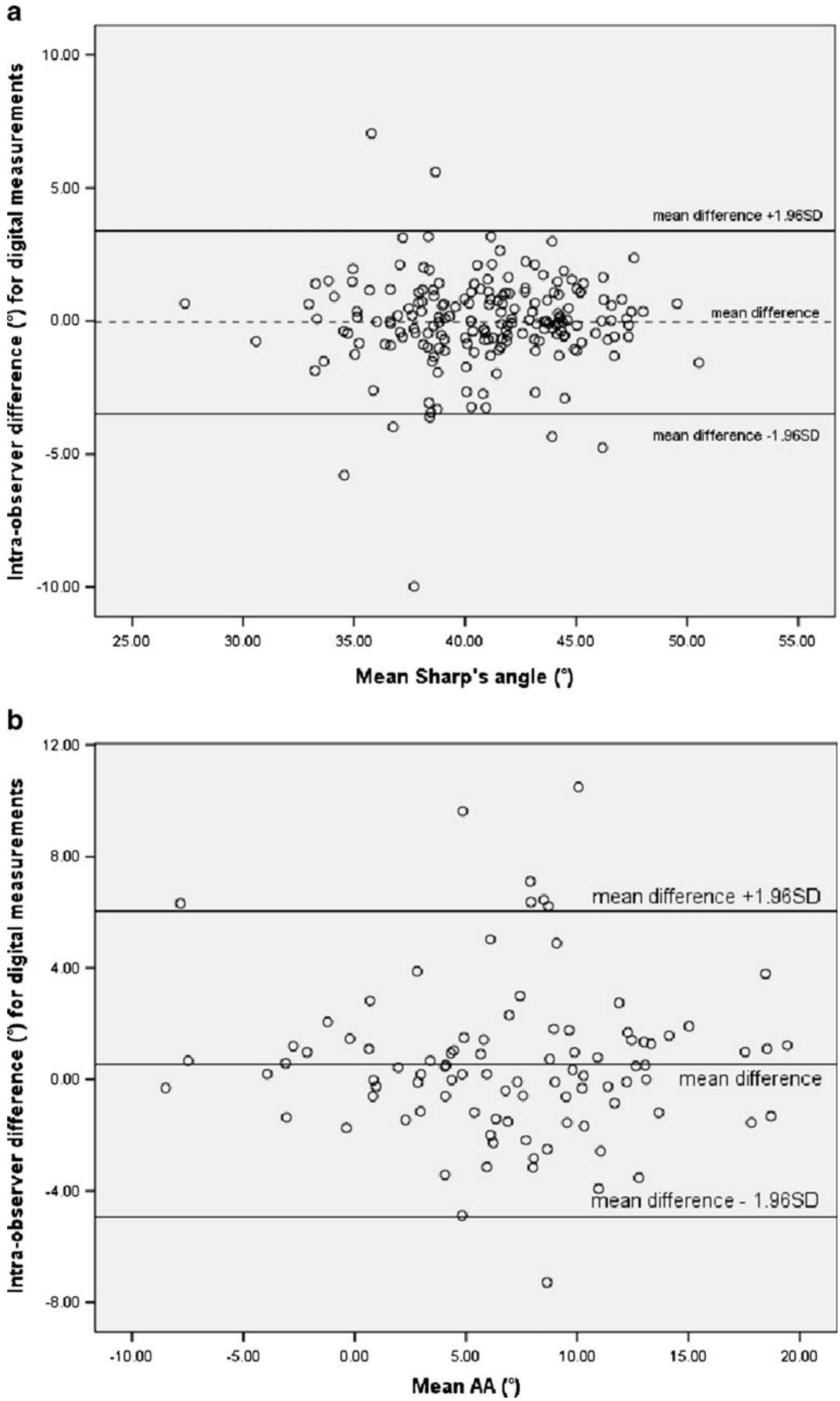

Fig. 4 Bland-Altman plot for digital intra-observer measurements for Sharp's angle (**a**) and acetabular roof angle of Tönnis (AA) (**b**)

compared to our findings of 3.0–3.4° for digital measurements and 2.6–3.7° for manual measurements. Other studies have also reported high values of SD and MDC for AA [10, 28].

Wiberg's CE angle and FHEI describe the position of the femoral head in the acetabular cavity (Fig. 1b). Wiberg's CE angle has been validated in several previous studies [9–12, 28, 29], with the reported inter-observer ICC values ranging from 0.73 to 0.92 as compared 0.84 to 0.91 in our study. Both Troelsen [11] and Lequesne [28] report on relatively wide within-observer variation, with SDs from 2.6 to 3.8° and 1.5 to 2.6°, respectively, as compared to ours (1.9–2.3° for digital measurements and 1.5–4.1° for manual measurements). For FHEI, there are, in principle, two different measurement methods. We used the method described by Heyman and Herndon [8] with a medial reference line drawn at the medial margin of the femoral head as compared to a second method using a line running through the most medial aspect of the joint space [30]. For both methods a cut-off value of less than 75% is considered pathological. Obviously FHEI obtained by the two methods will differ significantly in cases of subluxated or luxated femoral heads, underscoring the importance of standardization of measurements. In the current patient population no cases of (sub)luxations were identified. The reported levels of inter-observer variation for FHEI, using the ICC, range from 0.80 to 0.91. In comparison, we report ICCs of 0.82 to 0.84 and more importantly, 95% limits of agreements of less than 10% of the absolute values, which is considered acceptable for use in clinical studies.

Only a few studies have addressed the reproducibility of the maximum teardrop width, JSW, and ATD. In a study of 20 radiographs, Pedersen and colleagues found low levels of intra- and inter-observer reliability for both the maximum teardrop width and ATD [12], which are in good agreement with our results. JSW was validated by Lequesne et al., reporting SD of 0.4–0.5 mm and ICCs of 0.77–0.90 for intra-observer measurements, compared to SD 0.59–1.18 (digital) and 0.42–0.93 (manual) in our study.

In conclusion, large measurement variations are seen across different radiological measurements relevant for hip dysplasia at skeletal maturity, and this should be taken into account when evaluating dysplastic hips in clinical practice. Similar findings are found for both digital and manual measurement techniques.

Acknowledgments We thank radiographer Sigrun Tufta for performing all the high quality radiographs and dr Martin Biermann for all help and advice in the process of converting the pelvic radiographs to DICOM files. The study has received funding from University of Bergen and Arthritis Research UK (grant Ref 18196). Two of the authors (I.Ø.E., L.B.L.) are supported financially by Western Norway Regional Health Authority and one author (T.G.L.) by the Frank Mohn Foundation.

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

References

1. Engesæter IØ, Lehmann T, Laborie LB, Lie SA, Rosendahl K, Engesæter LB. Total hip replacement in young adults with hip dysplasia: age at diagnosis, previous treatment, quality of life, and validation of diagnoses reported to the Norwegian Arthroplasty Register between 1987 and 2007. Acta Orthop. 2011;82(2):149–154.
2. Dezateux C, Rosendahl K. Developmental dysplasia of the hip. Lancet. 2007;369(9572):1541–52.
3. Eastwood DM. Neonatal hip screening. Lancet. 2003;361 (9357):595–7.
4. Wiberg G. Studies on dysplastic acetabula and congenital subluxation of the hip joint. Acta Orthop Scand Suppl. 1939;58:1–132.
5. Sharp IK. Acetabular dysplasia. J Bone Jt Surg. 1961;43B (2):268–72.
6. Cooperman DR, Wallensten R, Stulberg SD. Acetabular dysplasia in the adult. Clin Orthop Relat Res. 1983;175:79–85.
7. Stulberg SD, Harris WH. Acetabular dysplasia and development of osteoarthritis of the hip. In: Harris WH, editor. The hip proceedings of the Second Open Scientific Meeting of the Hip Society. St Louis: Mosby; 1974. p. 82–93.
8. Heyman CH, Herndon CH. Legg-Perthes disease; a method for the measurement of the roentgenographic result. J Bone Jt Surg Am. 1950;32(A:4):767–78.
9. 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(856):331–4.
10. Mast NH, Impellizzeri F, Keller S, Leunig M. Reliability and agreement of measures used in radiographic evaluation of the adult hip. Clin Orthop Relat Res. 2011;469(1):188–99.
11. Troelsen A, Romer L, Kring S, Elmengaard B, Soballe K. Assessment of hip dysplasia and osteoarthritis: variability of different methods. Acta Radiol. 2010;51(2):187–93.
12. Pedersen DR, Lamb CA, Dolan LA, Ralston HM, Weinstein SL, Morcuende JA. Radiographic measurements in developmental dysplasia of the hip: reliability and validity of a digitizing program. J Pediatr Orthop. 2004;24(2):156–60.
13. Tönnis D. Normal values of the hip joint for the evaluation of X-rays in children and adults. Clin Orthop Relat Res. 1976;119:39–47.
14. Tönnis D, Brunken D. Differentiation of normal and pathological acetabular roof angle in the diagnosis of hip dysplasia. Evaluation of 2294 acetabular roof angles of hip joints in children. Arch Orthop Unfallchir. 1968;64(3):197–228.
15. Tönnis D, Legal H, Graf R. Congenital dysplasia and dislocation of the hip in children and adults. New York: Springer-Verlag; 1987.
16. Ogata S, Moriya H, Tsuchiya K, Akita T, Kamegaya M, Someya M. Acetabular cover in congenital dislocation of the hip. J Bone Jt Surg Br. 1990;72(2):190–6.
17. Albinana J, Morcuende JA, Weinstein SL. The teardrop in congenital dislocation of the hip diagnosed late. A quantitative study. J Bone Jt Surg Am. 1996;78(7):1048–55.
18. Langenskiold A, Salenius P. Epiphyseodesis of the greater trochanter. Acta Orthop Scand. 1967;38(2):199–219.
19. Jacobsen S, Sonne-Holm S, Soballe K, Gebuhr P, Lund B. Joint space width in dysplasia of the hip: a case-control study of 81 adults followed for ten years. J Bone Jt Surg Br. 2005;87(4):471–7.
20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10.
21. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res. 1999;8(2):135–60.

22. Bland JM, Altman DG. Agreement between methods of measurement with multiple observations per individual. J Biopharm Stat. 2007;17(4):571–82.
23. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996;1(1):30–46.
24. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59(10):1033–9.
25. Petrie A. Statistics in orthopaedic papers. J Bone Jt Surg Br. 2006;88(9):1121–36.
26. Bland JM, Altman DG. A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement. Comput Biol Med. 1990;20(5):337–40.
27. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1991.
28. Lequesne M, Malghem J, Dion E. The normal hip joint space: variations in width, shape, and architecture on 223 pelvic radiographs. Ann Rheum Dis. 2004;63(9):1145–51.
29. Tannast M, Mistry S, Steppacher SD, Reichenbach S, Langlotz F, Siebenrock KA, et al. Radiographic analysis of femoroacetabular impingement with Hip2Norm—reliable and validated. J Orthop Res. 2008;26(9):1199–205.
30. Delaunay S, Dussault RG, Kaplan PA, Alford BA. Radiographic measurements of dysplastic adult hips. Skelet Radiol. 1997;26(2):75–81.