EXTENDED REPORT

Reproducibility and inter-reader agreement of a scoring system for ultrasound evaluation of hip osteoarthritis

E Qvistgaard, S Torp-Pedersen, R Christensen, H Bliddal

Ann Rheum Dis 2006;**65**:1613–1619. doi: 10.1136/ard.2005.050690

See end of article for authors' affiliations

Correspondence to: Professor H Bliddal, The Parker Institute, Frederiksberg Hospital, DK 2000 Copenhagen F, Denmark; hb@fh.hosp.dk

Accepted 7 June 2005
Published Online First
25 May 2006

Objective: To evaluate the intra-reader and inter-reader agreements of ultrasonographic assessments of hip joints in patients with hip osteoarthritis.

Design: Ultrasonography was performed on 100 patients with hip osteoarthritis at 14 MHz using a 8–15 MHz linear probe. Dynamic sweeps of the hip and representative still images were used for the analysis. A semiquantitative grading score was introduced in the evaluation of the ultrasound pictures and compared with an overall ultrasound evaluation. The evaluation was performed by a specialist in ultrasonography and a rheumatologist trained in musculoskeletal ultrasound examination. Clinical pain assessment and joint aspiration were obtained in parallel with the ultrasonography.

Results: Intraobserver agreement represented by intraclass correlation coefficients (ICC) (exact agreement in percentage; unweighted κ values) showed good to excellent correlation, 0.8 with regard to the osteophyte score, 0.78 with regard to the femoral head score, 0.71 with regard to the fluid score and 0.69 with regard to the synovial profile score. Interobserver agreement was fair to good with corresponding ICC 0.65, 0.63, 0.45 and 0.6, respectively. In comparison, the ICC for the global osteoarthritis and synovial assessments were 0.7 and 0.72, respectively, for the intraobserver rating and 0.56 and 0.58, respectively, for the interobserver rating.

Conclusions: This study suggests that ultrasound is a reproducible method for the assessment of changes in the osseous surface and synovium-related inflammation. The semiquantitative scoring system presented seemed to match the global assessment of a trained ultrasound investigator and might be used by less-trained investigators.

The practical usage of ultrasound for examination of the hip joint has increased during the past decade.¹ Ultrasound is harmless to the patients; it does not entail ionising radiation and may be readily used by the clinician in an outpatient setting.² Ultrasound allows for a real-time evaluation of the joint and may also be used for verifying the placement of injections.³

Drawbacks are the subjectivity of the ultrasound evaluation and the lack of standards for the procedures and the diagnostic terminology. Despite the fast-growing availability and interest in musculoskeletal ultrasound,⁴ there is a shortage of studies dedicated to osteoarthritis of the hip and hence, a lack of standardisation. Ultrasound assessment of the hip has until now focused on the presence of effusion or synovitis by assessing the collum–capsule distance (CCD). Most studies have been carried out on paediatric patients, and the articles dealing with measurement of CCD in adults are based, by and large, on inflammatory rheumatic diseases.^{5–7} Even though measurements of well-defined distances have been shown to be reproducible at all levels of sonographic experience,⁸ the overall evaluation of a region is still considered to be highly operator dependent.

This study was undertaken to evaluate a scoring system for qualitative subclassification of ultrasound of the hip joint in osteoarthritis. Our primary goal was to investigate the repeatability and reproducibility of four specifically defined ultrasound parameters for hip osteoarthritis, as well as the global ultrasound assessments of bone and soft-tissue conditions. Our secondary goal was to investigate the clinical relevance of the chosen parameters compared with the radiographic findings.

MATERIALS AND METHODS

Patients

One hundred consecutive patients who, according to the American College of Rheumatology criteria,⁹ had radiographically verified hip osteoarthritis were included. In case of bilateral affection, the hip that caused most pain on the day of examination was chosen for imaging. The mean (standard deviation (SD)) age was 66 (12) years, range 28–88 years; 64% of the patients were female. Radiographic manifestations were representative of both light (Kellgren I+II, 57%) and severe (Kellgren III+IV, 43%) osteoarthritis. Clinical data were obtained by questionnaires stating the patient's "pain on walking" and "pain at rest" on a 100-mm Visual Analogue Scale (VAS).

Ultrasound examination

The patient was supine with the hip in neutral position. If this position was uncomfortable for the patient, slight flexion in the hip was obtained with a pillow behind the knees. The ultrasound scanning was performed with an Acuson Sequoia (Mountainview, California, USA) using a 5-cm linear (8–15 MHz) probe with a 14 MHz centre frequency (on one patient the joint was located so deeply that a 4 MHz curved probe was necessary). Both depth and focus of the image were adjusted for the position of the hip joint. "Chirp-coded excitation", which uses coherent pulse formation (control of the amplitude and phase of the transmitted wave form and

Abbreviations: CCD, collum–capsule distance; ICC, intraclass correlation coefficient; VAS, Visual Analogue Scale

Figure 1 Examples of osteophyte score. (A) Score 0, no visible osteophyte. (B) Score 1, irregularity on the cartilage–bone transition is just visible (arrow). (C) Score 2, well-defined osteophytes, shelf formation or irregularities on the femoral neck. There is one large osteophyte (f) with shelf formation (arrows), which is an ultrasound discontinuity between the distal border of the osteophyte and the femoral neck. (D) Score 3, involvement of the whole femoral neck including shelf formation. There is one large osteophyte (f) with shelf formation (horizontal arrows). The femoral head is seen between the two vertical arrows.

single-pulse capabilities) was applied to obtain deep penetration with this high-frequency transducer.

The joint was scanned in anterior longitudinal and transverse planes. The longitudinal plane was slightly angled to the sagittal plane and aligned with the axis of the femoral neck. Representative transverse and longitudinal images, as well as a 4-s (38 frames) live clip in the longitudinal plane, sweeping the joint from medial to lateral, were stored digitally in DICOM format. Patients included in the project underwent an ultrasound-guided aspiration in the target hip in continuation of the scanning, immediately after which they received intra-articular treatment.

Puncture

Non-touch technique was applied. With the patient in the supine position and after triple skin disinfection, a needle (gauge 21, 0.8×80 mm) was inserted inferiorly 8–10 cm under the inguinal ligament towards the anterior or inferior capsule below the femoral head with the free-hand technique. Guided by ultrasound, the needle was traced from 1 cm below the skin surface all the way to the joint. Joint fluid was aspirated if present.

Ultrasound image analysis

The live clips were exported and randomly numbered in a DICOM file. Each investigator chose the most representative image on the live clip and registered the frame number, which corresponds to taking a representative still image during a scanning session. For each patient, one selected image was scored for osteophytes, condition of the femoral head, presence of joint effusion and synovial profile using a semiquantitative scoring system.

The parameters were defined as follows:

- Osteophyte score (fig 1) described the femoral osteophytes:

Figure 2 Examples of femoral head score. (A) Score 0, round femoral head. (B) Score 1, slightly flattened femoral head. (C) Score 2, very flattened femoral head (f). (D) Score 3, no obvious contour—the femoral head (f) cannot be identified.

- 0, no occurrence;
- 1, slight degree (irregularity on the cartilage–bone transition is just visible);
- 2, medium degree (well-defined osteophytes, shelf formation or irregularities on the femoral neck);
- 3, severe degree (involvement of the whole femoral neck including shelf formation).
- Femoral head score (fig 2) described the curvature of the visible part:
 - 0, round;
 - 1, slightly flattened (still visible curvature but with an abnormally large radius);
 - 2, very flattened (no visible curvature of the caput);
 - 3, no obvious contour (the femoral head cannot be defined for osteophytes/erosions).
- Synovial profile (fig 3) was defined from the course of the anterior surface of the capsule on the anterior surface of the femoral neck (technically including effusion, synovium and capsule):
 - 0, concave (follows the bone surface);
 - 1, flat;
 - 2, convex.
- Joint effusion (fig 4) was defined as a hypoechoic coherent region present inside the synovium delimitation:
 - 0, none;
 - 1, perhaps present;
 - 2, present.

The overall assessment of the hip by the investigator was also standardised semiquantitatively as follows:

Figure 3. Examples of joint effusion score. (A) Score 0, no fluid. The synovial space is uniformly hypoechoic, without areas suggesting the presence of fluid. (B) Score 1, fluid is perhaps present (arrow). (C) Score 2, fluid is present (}).

- Global ultrasound evaluation of osteoarthritis:

- 0, normal;
- 1, slight;
- 2, moderate;
- 3, severe.

- Global ultrasound evaluation of synovitis:

- 0, none;
- 1, moderate;
- 2, severe.

Study design

Two independent investigators examined the images. One was an ultrasound expert with extensive experience in musculoskeletal ultrasound (A; ST-P) and the other a rheumatologist with 4 years, training in musculoskeletal ultrasound (B; EQ). Investigator B performed all ultrasound examinations.

An ultrasound-guided aspiration in the target hip was performed at the end of the scanning.

The performance of the two investigators was examined with both intraobserver and interobserver variation as follows:

1. Which of the 38 consecutive frames of the clip was chosen as being representative of the joint? (time 1)
2. Based on the frames selected by investigator B, investigator A scored these still images twice (intraobserver; times 2 and 3), whereas investigator B scored the still images once (time 2), independent of the frame-selection performance (interobserver).

There was an interval of at least 6 months between the actual scanning and the evaluation of live clips (time 1). There was a 4-week interval between all time slots.

Statistical analyses

Intraobserver and interobserver agreements were estimated using interclass correlation coefficients (ICC): ICC (1;1); (one-way random single measure) for intra-rater analysis and ICC (2;1) (two-way mixed model, absolute agreement definition) for inter-rater analysis by means of the statistical software SPSS V.11. Unweighted κ values and overall agreement (defined as the percentage of observed exact agreements) were calculated. The reliability is regarded as excellent if $ICC > 0.75$, fair to good if $0.4 < ICC < 0.75$, and poor if $ICC < 0.4$.¹⁰ κ is defined as being almost perfect if > 0.81 , substantial if between 0.61 and 0.80, moderate if between 0.41 and 0.60, fair if between 0.21 and 0.40, slight if between 0.20 and 0, and poor if < 0 .¹¹

The relationship between the results of the two investigators was reported using the Bland–Altman plot, a graphical illustration illustrating repeatability, and a corresponding two-way table.

Associations between the ultrasound scores, radiographic scores (Kellgren) and clinical outcomes were evaluated by univariate (linear) regression, complemented by multivariate regression analyses of clinical, ultrasound and radiographic data, using the MAXR (SAS) procedure. All regression analyses were carried out using the SAS statistical package, V.8.

RESULTS

Observer agreement

A total of 100 hips were analysed by investigator A. Table 1 summarises the analysis of intraobserver agreement. The ICC for the examined parameters showed good to excellent correlation (0.69–0.80) with the matching κ varying from moderate to substantial (0.55–0.75). The overall agreement was high (74–87%). In all three ways of considering the agreement, the lowest values were found in the assessments of the synovium profile. Interobserver agreement between ultrasound investigators was also determined for 100 osteoarthritis hips and showed a fair to good ICC (0.45–0.65), whereas the

Figure 4. Examples of synovial profile score. (A) Score 0, concave. The anterior border of the capsule is convex. (B) Score 1, flat. The anterior surface of the capsule is flat. (C) Score 2, convex. The anterior surface of the capsule is convex. In all three images the synovium is thickened.

Table 1 Intraobserver and interobserver agreement between the two ultrasound investigators as measured by three different statistical methods*

Score	Intraobserver			Interobserver		
	ICC	κ	Overall agreement (%)	ICC	κ	Overall agreement (%)
Osteophytes	0.80	0.75	87	0.65	0.49	69
Femoral head	0.78	0.69	79	0.63	0.35	54
Joint effusion	0.71	0.58	75	0.45	0.38	63
Synovial profile	0.69	0.55	74	0.60	0.43	65
Global US evaluation of osteoarthritis	0.70	0.60	79	0.56	0.30	54
Global US evaluation of synovitis	0.72	0.63	81	0.58	0.37	62

ICC, intraclass correlation coefficient; US, ultrasound.

*Intraclass correlation coefficient; unweighted κ and percentages.

corresponding κ was fair to moderate (0.30–0.49), with moderate overall agreement (54–69%).

As shown by the Bland–Altman plots (fig 5) and table 2, there were no systematic differences in the scoring of the two investigators.

Choice of representative frame

In >50% of the cases, the two investigators had their frame of choice placed within three frames of a possible 38. The most

divergent choice was 18 frames apart. Figure 6 shows the distribution of the frames.

Ultrasound versus radiography

Use of Spearman's r model indicated weak correlation coefficients between the Kellgren scores and the two osseous ultrasound scores. Kellgren versus osteophyte score: $r = 0.26$, $p = 0.017$; Kellgren versus femoral head score: $r = 0.24$, $p = 0.03$.

Figure 5 Adapted Bland–Altman plots illustrating test–retest variability of the four semiquantitative scorings by mapping the difference of paired variables versus their average. Each point on the graphs is a “flower” where the number of “petals” represents the number of patients exceeding one at that point.

Table 2 Distribution of the four semiquantitative scorings by the two investigators

		Osteophyte			
		Observer B			
Score		0	1	2	3
Observer A	0	2	1	0	0
	1	0	2	1	0
	2	3	12	20	5
	3	1	0	8	45

		Femoral head			
		Observer B			
Score		0	1	2	3
Observer A	0	9	5	0	0
	1	11	15	2	0
	2	2	13	26	1
	3	1	1	10	4

		Effusion		
		Observer B		
Score		0	1	2
Observer A	0	36	8	2
	1	16	22	7
	2	3	1	5

		Osteophyte		
		Observer B		
Score		0	1	2
Observer A	0	10	4	2
	1	8	15	9
	2	2	10	40

The values are reported as percentage of cases with same category, 1 category discrepancy, 2 categories discrepancy, etc.

Univariate regression (table 3) showed p values <0.001 for all the available parameters. The association between the patient's pain at rest and pain on walking assessments in mm VAS and the ultrasound scores was largest for global osteoarthritis and osteophytes, with significant associations also for the Kellgren score; pain at rest 57–59.5% and pain on walking 72.5–73.9%.

When testing whether more than one independent variable improved prediction of the dependent variable, a multivariate regression analysis procedure, with a pre-defined significance level ($\alpha \leq 0.1$), was applied to show independent variables of significance for pain. With regard to the dependent variable, VAS pain at rest, only two variables had mutually independent predictive properties: VAS activity ($p < 0.001$) and femoral head score ($p = 0.060$), with $R^2 = 74.3\%$.

With regard to the dependent variable, VAS pain on activity, four variables had mutually independent predictive properties: VAS pain at rest ($p < 0.001$), global osteoarthritis ($p = 0.001$), femoral head score ($p = 0.006$) and Kellgren score ($p = 0.008$), with $R^2 = 84.0\%$.

The outcome of the aspirations in the 100 patients contained fluid in 21 patients, with typical amounts of <1 ml (mean 0.82 ml, range 0.1–2.0 ml). Table 4 shows the distribution of effusion scores and aspirated fluid. No systematic differences were observed in the scoring of patients and there was no association between aspiration and fluid on ultrasound.

A retrospective multivariate regression analysis was applied to test whether the ultrasound measures of general osteoarthritis and global synovitis performed equally as the separate image analysis in relation to the patients' self-reported pain.

Figure 6 Diagram showing the concordance between investigators A and B in the selection of a representative frame out of a 38-image sweep of an osteoarthritis-affected hip.

According to the VAS pain on activity, a highly significant association with the US global evaluation ($R^2 = 0.73$, $p < 0.001$) and the ultrasound synovitis estimate by the reader was also of predictive importance for the VAS pain on activity ($R^2 = 0.59$, $p < 0.001$; table 3). The combination of these two measures gave significant additive information for VAS pain on activity as the ultrasound global evaluation ($p < 0.001$) and the synovitis score ($p = 0.035$) with a small increment in predictability ($R^2 = 0.74$).

DISCUSSION

Owing to the relatively young age of the ultrasound modality, great efforts are still made to achieve standardisation in musculoskeletal sonography as well from a reference point of view as to the actual observer reliability of targeted regions.^{12–13} Despite its importance for mobility, the hip joint still seems to be under-rated in this respect and to our knowledge, no standardised way of reporting ultrasound examinations of this joint has yet been developed for the evaluation of osteoarthritis.

From our experience with these examinations, we chose to test four ultrasound parameters covering the important aspects of hip osteoarthritis—that is, the bony and the synovial changes. In addition, the investigators were asked to give a global ultrasound evaluation, corresponding to an overall assessment. Different aspects of both the bony changes and the synovial evaluations showed a fair amount of consistency between the two evaluators. ICC was chosen as the primary outcome, owing to the ordinal character of our data and κ values, and absolute agreements were calculated as complementary information.

In clinical practice, the interpretation of which image is representative of the observed hip might be of great importance for the assessment. The distribution of the chosen frames in our study showed good concordance between the two observers when interpreting osteoarthritis. Secondary analysis of the outliers diminished the concerns of possible misinterpretations by showing widespread arthrotic changes of similar magnitude.

The intra-observer repeatability in our study was good to excellent, with no ICC < 0.69 and the κ results followed the same pattern. κ Values were comparable to those observed for the Kellgren score in radiological evaluation of the hip joint.¹⁴ Even though there were slightly weaker values for the assessment of soft tissue as opposed to that of bony structures, the overall ICC results of the partitioned scoring system showed stability comparable with global scores (ICC 0.70–0.72). The accuracy of CCD has been questioned when magnetic resonance imaging is used as the gold standard,⁵ and has therefore not been taken into account in this study.

Table 3 Univariate regression coefficients (β), presented in descending order, according to the percentage variation predicted ($R^2\%$) for the association of specialist-assessed osteoarthritis outcomes for pain at rest and pain on activity

	R^2 (%)	β	SE (β)	p Value
VAS; pain at rest (scale: 0–100 mm)				
Independent variable				
VAS pain on walking	73.8	0.55	0.03	<0.001
Global OA	59.5	9.71	0.81	<0.001
Osteophyte score	58.1	9.57	0.82	<0.001
Kellgren score	57.0	9.05	0.88	<0.001
Femoral head score	49.6	12.20	1.24	<0.001
Global synovitis	47.9	17.23	1.82	<0.001
Synovial profile	46.7	14.10	1.52	<0.001
Effusion	37.0	21.56	2.84	<0.001
VAS; pain on walking (scale: 0–100 mm)				
Independent variable				
Kellgren score	73.9	16.40	1.09	<0.001
Osteophyte score	73.8	16.82	1.01	<0.001
VAS pain at rest	73.8	1.34	0.08	<0.001
Global OA	72.5	16.71	1.04	<0.001
Global synovitis	59.3	29.92	2.50	<0.001
Synovial profile	58.3	24.56	2.10	<0.001
Femoral head score	58.1	20.59	1.77	<0.001
Effusion	42.4	35.99	4.24	<0.001

OA, osteoarthritis; VAS, Visual Analogue Scale.
Assessments conducted by investigator B.

According to our ultrasound examiner test, the global osteoarthritis score was the most closely related to the symptoms of the patients estimated by the retrospective test of the VAS activity, whereas the synovitis score had an independently significant, albeit small importance of its own (about 1% improved prediction of VAS activity), indicating a possible influence of inflammation on osteoarthritis of the hip. This result is in accordance with the notion of inflammation in osteoarthritis as indicated by ultrasound on examination of the knee.¹⁵

The interobserver ICCs were lower than their intra-observer equivalents, yet still within the range of acceptability (ICC 0.45–0.65; κ 0.35–0.49). A trend in favour of bony contours was found, similar to the intraobserver variation, but here too, the partitioned scoring systems were found to be in the same range as the global score (ICC 0.58–0.56).

The presence of effusion is of the greatest interest when trying to expose synovitis. However, the present study showed a low ICC of 0.45 (effusion). Some of the disagreement could be due to a difference in cut-off levels. The actual presence of joint exudates summed up (when dichotomising the fluid score into positive and non-positive) into positive predictive values of 0.44/0.5 and negative predictive values of 0.881/0.83 for examiners A and B, respectively. Several explanations may be offered in this connection: large viscosity of the fluid, occlusion of the rather thin needle (21 G) and uneven distribution of the fluid away from the site of puncture are just some of the factors against using the aspirated volume as gold standard. On the other hand, the capsule may be so tight that a rather limited effusion may result in high intra-articular pressure and patient discomfort.

In this regard, it is interesting to note that in only a few of our patients did we observe fluid accumulation after intra-articular injection of 3-ml volume.

Ultrasound may allow visualisation of even very small fluid accumulations, although proof of this assumption can only be obtained by aspiration, or if the amount of fluid leads to a displacement of fluid under applied pressure. As a parameter in a general score of the hip in osteoarthritis, fluid is of limited value, although in the final diagnosis of the ultrasound examination it cannot be disregarded.

Osteophytes on the acetabulum were left out in this study, as we assessed their presence to be of minor importance and only for the impairment of joint motion.

The clinical association of the various imaging parameters was tested by univariate linear regression, analysing what percentage (R^2) of our clinical outcomes (self-reported pain on VAS) could be explained by each chosen parameter.

The high significance of the ultrasonographic findings confirmed the overall relevance of our choice of parameters.

The smallest degree of predictability was observed for pain at rest, with only 58.1% explained statistically by the ultrasound osteophyte score. The radiographic score was in this matter equipotent ($R^2 = 57\%$). By contrast, the synovial parameters (shape and effusion) were both below $R^2 = 50\%$.

Pain on walking, however, was explained with as much as $R^2 = 73.8\%$ by the osteophyte score, which was similar to the result of the Kellgren score, $R^2 = 73.9\%$. Only the effusion score had no explanatory value.

We also wanted to show whether x ray and ultrasound were somehow associated and therefore possibly confounded. The stepwise multivariate regression analysis is designed to

Table 4 Distribution of effusion scores and aspirated fluid

Effusion score	Investigator A		Investigator B	
	n	Patients with aspirated fluid	n	Patients with aspirated fluid
0	46	7	55	4
1	45	10	31	10
2	9	4	14	7
Total	100	21	100	21

eliminate mutual predictive variables, thus pointing to independent variables representing a better prediction. With the natural exception of pain on walking, the second best predictive parameter of pain at rest was the femoral head score. Similarly, pain on walking was predicted by the Kellgren score, as well as the structural descriptive parameters of ultrasound—that is, the femoral head score and global osteoarthritis score (which was dependent on the osteophyte score more than on any other factor).

Radiographic and ultrasound scores presented independent factors in the stepwise, multivariate analysis, and the independence between radiographic and US scores in the assessment of osteoarthritis hips suggest that the two modalities register different characteristics of the disease, reflecting the difference in the perspective of the ultrasound findings.

In conclusion, this study suggests that ultrasound could be a reproducible method for the assessment of changes in the osseous surface and synovium-related inflammation. The semiquantitative scoring system presented seemed to match the global assessment of a trained ultrasound investigator and might be used by moderately trained investigators—for example, clinicians—but only after a proper introduction to the procedures involved in a systematic evaluation of osteoarthritis.

Ultrasound is less often used in the adult hip compared with other major joints such as the shoulder and ankle, largely because of the relative inaccessibility.¹⁶ Clinical examination is impaired for the same reasons, and it is our concern that the osteoarthritis hip may have a tendency to be underdiagnosed. Ultrasound should not be regarded as a substitute for radiographic assessment, but rather as a supplementary source of information.

Future studies are needed to clarify the validity of the score in clinical situations.

Authors' affiliations

E Qvistgaard, S Torp-Pedersen, R Christensen, H Bliddal, The Parker Institute, Frederiksberg Hospital, Copenhagen, Denmark

Funding: This study was supported by the Oak Foundation and the Erna Hamilton Foundation.

Competing interests: None declared.

REFERENCES

- 1 **Grassi W**, Cervini C. Ultrasonography in rheumatology: an evolving technique. *Ann Rheum Dis* 1998;**57**:268–71.
- 2 **Manger B**, Kalden JR. Joint and connective tissue ultrasonography—a rheumatologic bedside procedure? A German experience. *Arthritis Rheum* 1995;**38**:736–42.
- 3 **Qvistgaard E**, Kristoffersen H, Terslev L, Danneskiold-Samsøe B, Torp-Pedersen S, Bliddal H. Guidance by ultrasound of intra-articular injections in the knee and hip joints. *Osteoarthritis Cartilage* 2001;**9**:512–17.
- 4 **Wakefield RJ**, Goh E, Conaghan PG, Karim Z, Emery P. Musculoskeletal ultrasonography in Europe: results of a rheumatologist-based survey at a EULAR meeting. *Rheumatology (Oxford)* 2003;**42**:1251–3.
- 5 **Soini I**, Kotaniemi A, Kautiainen H, Kauppi M. US assessment of hip joint synovitis in rheumatic diseases. A comparison with MR imaging. *Acta Radiol* 2003;**44**:72–8.
- 6 **Koski JM**, Anttila P, Hamalainen M, Isomaki H. Hip joint ultrasonography: correlation with intra-articular effusion and synovitis. *Br J Rheumatol* 1990;**29**:189–92.
- 7 **Sada PN**, Rajan P, Jeyaseelan L, Washburn MC. Standards for ultrasonographic measurements of the hip joint in Indian adults. *Skeletal Radiol* 1994;**23**:111–12.
- 8 **Balint PV**, Sturrock RD. Intraobserver repeatability and interobserver reproducibility in musculoskeletal ultrasound imaging measurements. *Clin Exp Rheumatol* 2001;**19**:89–92.
- 9 **Altman R**, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. *Arthritis Rheum* 1991;**34**:505–14.
- 10 **Fleiss JL**. *The design and analysis of clinical experiments*. New York: John Wiley & Sons, 1986.
- 11 **Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–174.
- 12 **Naredo E**, Moller I, Moragues C, de Agustin JJ, Scheel AK, Grassi W, et al. Inter-observer reliability in musculoskeletal ultrasonography: results from a "Teach-the-Teachers" rheumatologist course. *Ann Rheum Dis* 2006;**65**:14–19.
- 13 **Schmidt WA**, Schmidt H, Schicke B, Gromnica-Ihle E. Standard reference values for musculoskeletal ultrasonography. *Ann Rheum Dis* 2004;**63**:988–94.
- 14 **Reijman M**, Hazes JM, Pols HA, Bernsen RM, Koes BW, Bierma-Zeinstra SM. Validity and reliability of three definitions of hip osteoarthritis: cross sectional and longitudinal approach. *Ann Rheum Dis* 2004;**63**:226–32.
- 15 **D'Agostino MA**, Conaghan P, Le Bars M, Baron G, Grassi W, Martin-Mola E, et al. EULAR report on the use of ultrasonography in painful knee osteoarthritis part 1: prevalence of inflammation in osteoarthritis. *Ann Rheum Dis* 2005;**64**:1703–9.
- 16 **Cardinal E**, Bureau N. Hanche de l'adulte. In: Peetrons P, Chhem R, eds. *Atlas d'échographie du système locomoteur*. Tome II. Montpellier: Sauramp médical, 2000:139–53.