


BMJ Open Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis: an inter-rater and intra-rater reliability study

Stine Clausen ^{1,2} Søren Kjær,³ Ulrich Fredberg,^{3,4} Lene Terslev,^{5,6} Jan Hartvigsen,^{1,7} Bodil Arnbak^{1,2}

To cite: Clausen S, Kjær S, Fredberg U, *et al.* Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis: an inter-rater and intra-rater reliability study. *BMJ Open* 2020;**10**:e038643. doi:10.1136/bmjopen-2020-038643

► Prepublication history and additional material for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-038643>).

Received 19 March 2020
Revised 05 October 2020
Accepted 17 October 2020



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Stine Clausen;
sclausen@health.sdu.dk

ABSTRACT

Objectives The objectives of this study were to assess (1) inter-rater and intrarater reliability of ultrasound imaging in patients with hip osteoarthritis, and (2) agreement between ultrasound and X-ray findings of hip osteoarthritis using validated Outcome Measures in Rheumatology ultrasound definitions for pathology.

Design An inter-rater and intrarater reliability study.

Setting A single-centre study conducted at a regional hospital.

Participants 50 patients >39 years of age referred for radiography due to hip pain and suspected hip osteoarthritis were included. Exclusion criteria were previous hip surgery in the painful hip, suspected fracture or malignant changes in the hip.

Intervention Bilateral ultrasound examinations (n=92) were performed continuously by two experienced operators blinded to clinical information and other imaging findings. After 4–6 weeks, one operator reassessed the images. X-rays were assessed by a third imaging specialist.

Primary and secondary outcome measures Inter-rater and intrarater reliability and agreement between ultrasound imaging and X-ray were assessed using Cohen's ordinal kappa statistics for binary categorical variables and weighted kappa for ordered categorical variables.

Results Kappa values (κ) for inter-rater reliability were 0.9 and 0.8 for hip effusion/synovitis and osteoarthritis grading, respectively. For acetabular and femoral osteophytes, femoral cartilage changes and labrum changes κ ranged from 0.4 to 0.7. Intrarater reliability had κ equal or higher compared with inter-rater reliability. Agreement between ultrasound and X-ray findings ranged from $\kappa=0.2$ to $\kappa=0.5$.

Conclusion This study demonstrated substantial to almost perfect reliability on the most common ultrasound findings related to hip osteoarthritis and osteoarthritis grading. Agreement on the grade of osteoarthritis between ultrasound and X-ray was moderate. Overall, these results support ultrasound imaging as a reliable tool in the assessment of hip osteoarthritis.

INTRODUCTION

Osteoarthritis (OA) is characterised by progressive destruction of articular cartilage, changes in bone tissue, osteophyte formation

Strengths and limitations of this study

- Using Outcome Measures in Rheumatology validated ultrasound definitions makes the results applicable for other clinicians.
- Including the acquisition technique in the assessment of inter-rater reliability is of great importance due to the dynamic nature of ultrasound.
- The participants represent a broad spectrum of patients with hip pain and suspected hip osteoarthritis, making the results transferable to other clinical settings.
- Intrarater reliability was investigated by reassessing existing still images from the ultrasound examination, and therefore did not include findings that required dynamic examination.
- X-rays were recorded according to Department guidelines, and variation in acquisition procedures might have affected the findings.

and joint inflammation resulting in loss of normal joint function.¹ The pathophysiological changes can be visualised with a broad spectrum of imaging modalities. Plain X-ray can detect structural changes in bones but give little information about soft tissue and inflammatory changes, whereas ultrasound and MRI help you visualise both structural changes in the articular bone and inflammatory changes in soft tissue around the joint.² MRI has the advantage of revealing intra-articular structures better than all other modalities³; however, access to MRI can be restricted due to its high expense and limited availability. Ultrasound, on the other hand, is relatively inexpensive and accessible and allows for dynamic examination, assessment of Doppler activity and clinician–patient interaction during examination.⁴ Therefore, ultrasound imaging is increasingly used in research to provide insight into the pathophysiology of OA. Ultrasound has limitations in showing intra-articular structures and

pathology within bones and is criticised for the lack of validation and high degree of operator dependence.

The most frequent findings on diagnostic ultrasound in hip OA include joint effusion, synovial thickening, cartilage destruction including degeneration of labrum, subchondral cystic lesions and osteophytes.^{5–6} Iliopsoas bursitis rarely occurs, but prevalence increases at more advanced stages of OA.⁷ Acceptable reliability of ultrasound-specific lesions provides the foundation for diagnostic or epidemiological studies using ultrasound imaging, but only a few previous studies have investigated the reliability of hip ultrasound in OA and they have mostly investigated individual findings and only by interpreting the same images.⁶ Therefore, studies that include the differences in acquisition of images between the two operators in the assessment of reliability on ultrasound findings in people with hip OA are needed.

The primary objective of this study was to assess the inter-rater and intra-rater reliability of ultrasound findings in patients with hip OA. The trochanter region was also examined in order to investigate for bursitis, as patients with hip pain often complain of pain in this location.⁸ The secondary aim was to assess agreement between ultrasound and radiological findings related to hip OA.

MATERIALS AND METHODS

Guidelines for Reporting Reliability and Agreement Studies⁹ were used.

Data collection occurred from December 2018 until April 2019. Patients older than 39 years, referred to the Department of Radiology, Silkeborg Hospital, for radiography due to hip pain and suspected hip OA were included. Patients were excluded if they had previous hip surgery in the painful hip, suspected fracture or malignant changes in the hip or if the patient did not read and speak Danish. The sample size was chosen based on literature recommendations.¹⁰

All participants completed an electronic questionnaire containing demographic data, and the Danish version of the Hip disability and Osteoarthritis Outcome Score (HOOS) questionnaire,¹¹ which assesses hip pain and function.

Ultrasound imaging

Bilateral ultrasound examination of hips and trochanter regions, regardless of unilateral pain, was performed using a high-end ultrasound device (HI VISION Ascendus, Hitachi Medical Systems, Steinhausen, Swiss) with an 18.5 MHz linear transducer (central frequency of 9 MHz) and the possibility of trapezoidal imaging. Predefined settings were used, with individual adjustment of the overall gain, depth and focus. The examinations were performed continuously, based on a protocol defined by the European League Against Rheumatism.¹² Patients were examined supine with straight legs and 15°–20° of external rotation of the hip. The trochanter region was examined with the patient lying on the opposite side with

15°–20° of flexion in the hip and knee. Study time for each hip including collection of data was 10–15 min for each examiner.

The ultrasound operators were a chiropractor and a rheumatologist. Both had 10–15 years of experience using musculoskeletal ultrasound (ultrasound qualification equivalent to European Federation of Societies for Ultrasound in Medicine and Biology level 2).¹³ They were blinded to the patient's clinical information and to each other's findings.

Prior to inclusion of participants, the two operators performed consensus sessions examining 10 patients, who would have met the inclusion criteria.

The Outcome Measures in Rheumatology (OMERACT) ultrasound definitions for osteophytes, cartilage, effusion and synovial hypertrophy were used.¹⁴ Labrum changes were assessed according to Martinoli *et al*¹⁵ and graded with our own staging as none (homogeneous echogenicity), mild (heterogeneous echogenicity and labrum poorly defined), moderate (definite pathology such as tears or cysts) or severe (pathology or degeneration to a degree where the labrum could not be defined). Femoral head deformation was assessed and rated semiquantitatively (none, mild, moderate or severe) according to a scoring system for the shape of the femoral head described by Qvistgaard *et al*.¹⁶ Trochanter and iliopsoas bursitis were scored dichotomously according to whether there was effusion in the bursa (present/absent).¹⁷ The ultrasound findings assessed, grading systems and definitions are listed below and described in online supplemental file 1. Image examples are illustrated in online supplemental file 2.

Ultrasound examination of anterior hip

Osteophytes on the anterior femur and acetabular rim and the femoral cartilage changes on the anterior articular surface of the femoral head were assessed. A measurement of the cartilage thickness was made as close

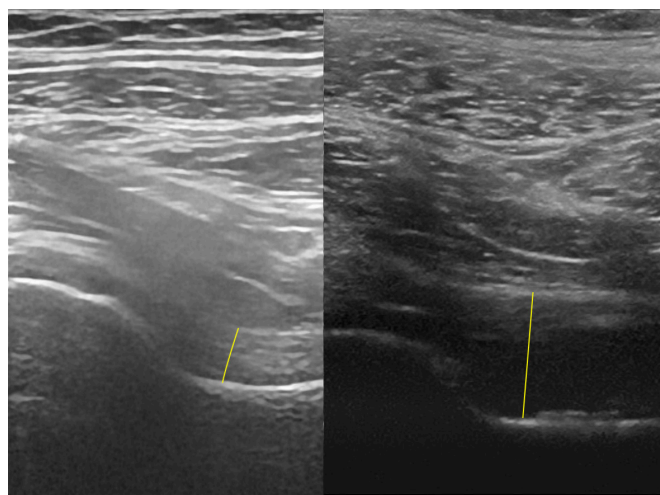


Figure 1 Anterior hip joint recess, longitudinal scan. The yellow line marks the bone-capsule distance. The hip on the right has effusion/synovitis.

Table 1 Prevalence and mean measure of ultrasound findings in 50 participants (92 hips) examined by two operators blinded to each other's findings

Ultrasound finding	Prevalence Operator A	Prevalence Operator B
Joint recess profile (%)		
Concave	71 (77)	73 (79)
Straight	12 (13)	10 (11)
Convex	9 (10)	9 (10)
Not accessible	0	0
Overall joint effusion/synovitis (%)		
Absent	74 (80)	73 (79)
Present	18 (20)	19 (21)
Not accessible	0	0
Femoral osteophytes (%)		
None	23 (25)	18 (20)
Mild	41 (45)	43 (47)
Moderate	19 (21)	22 (24)
Severe	9 (10)	9 (10)
Not accessible	0	0
Acetabular osteophytes (%)		
None	38 (41)	40 (44)
Mild	31 (34)	38 (42)
Moderate	18 (20)	11 (12)
Severe	5 (5)	1 (1)
Not accessible	0	2 (1)
Femoral head deformation (%)		
None	57 (62)	49 (53)
Mild	23 (25)	29 (32)
Moderate	8 (9)	10 (11)
Severe	4 (4)	4 (4)
Not accessible	0	0
Femoral cartilage changes (%)		
None	33 (36)	25 (27)
Mild	31 (34)	36 (39)
Moderate	23 (25)	27 (29)
Severe	5 (5)	4 (4)
Not accessible	0	0
Labrum changes (%)		
None	17 (18)	9 (10)
Mild	21 (23)	18 (20)
Moderate	17 (18)	19 (21)
Severe	32 (35)	33 (36)
Not accessible	5* (5)	13* (14)
Iliopsoas bursitis (%)		
Absent	89 (97)	88 (96)
Present	3 (3)	4 (4)
Not accessible	0	0

Continued

Table 1 Continued

Ultrasound finding	Prevalence Operator A	Prevalence Operator B
Trochanter bursitis (%)		
Absent	90 (98)	86 (94)
Present	2 (2)	5 (6)
Not accessible	0	0
OA grading (%)		
None	32 (35)	27 (29)
Mild	27 (29)	35 (38)
Moderate	25 (27)	22 (24)
Severe	8 (9)	8 (9)
Not accessible	0	0
	Mean distance Operator A	Mean distance Operator B
Bone-capsule distance (including capsule) (mm)	6.7 (SD 2.5)	6.4 (SD 2.2)
Bone-capsule distance (excluding capsule) (mm)	5.8 (SD 2.3)	5.6 (SD 1.9)
Femoral cartilage (mm)	0.8 (SD 0.3)	0.9 (SD 0.3)

*Labrum was not rated if operators did not find it possible to visualise satisfactorily.
OA, osteoarthritis.

to the labrum as possible. If the cartilage was very irregular, it was noted that a trustworthy measurement was not possible to obtain.

Effusion/synovitis was assessed in three different ways: (1) Measuring the bone-capsule distance (BCD) in the anterior joint recess in the longitudinal plane of the femoral neck (figure 1). We measured BCD from the cortical surface of the femoral neck to both the inner and the outer edges of the joint capsule, the latter combining joint fluid and synovium/capsule. A BCD increase of 7 mm or more (inner edge of joint capsule) or a bilateral difference of 1 mm indicates effusion according to Koski criteria.¹⁸ (2) A categorical assessment of the course of the anterior joint recess along the anterior surface of the femoral neck. Presence of a straight or convex joint recess indicates effusion/synovitis.¹⁹ (3) An overall assessment of effusion/synovitis was performed based on the joint recess profile and BCD using Koski criteria,¹⁸ and the possible presence of hypochoic or anechoic fluid in the joint recess along the femoral neck, and recorded as present or absent.

At the end of the examination, the operator rated the degree of hip OA equivalent to Kellgren-Lawrence grading (KLG) system, however based on the ultrasound findings as none (normal findings, only small osteophytes or subtle changes in the cartilage), mild (mild but definite changes in the femoral cartilage, small osteophytes, possible labral degeneration) or moderate or severe—increasing graduation with progressive change. This OA

Table 2 Prevalence of osseous findings and of OA on ultrasound (US) and radiographs, respectively, in the 63* hips that were X-rayed

Findings rated on ultrasound and X-rays	Prevalence US Operator A	Prevalence US Operator B	Prevalence on radiographs
Femoral osteophytes (%)			
None	16 (25)	14 (22)	44 (70)
Mild	27 (43)	28 (44)	16 (25)
Moderate	15 (24)	15 (24)	2 (3)
Severe	5 (8)	6 (10)	1 (2)
Acetabular osteophytes (%)			
None	27 (43)	26 (42)	19 (30)
Mild	21 (33)	28 (45)	29 (46)
Moderate	12 (19)	7 (11)	13 (21)
Severe	3 (5)	1 (2)	2 (3)
Femoral head deformation (%)			
None	39 (62)	33 (52)	40 (63)
Mild	17 (27)	21 (33)	18 (29)
Moderate	5 (8)	7 (11)	2 (6)
Severe	2 (3)	2 (3)	1 (2)
OA grading† (%)			
KLG 0 and 1	23 (37)	20 (32)	27 (43)
KLG 2	18 (29)	21 (33)	17 (27)
KLG 3	17 (27)	17 (27)	15 (24)
KLG 4	5 (8)	5 (8)	4 (6)

*Because we followed the Department guidelines all participants had an X-ray of the painful hip, but only some had bilateral hip (pelvic AP) resulting in 63 hip X-rays.

†In ultrasound the degree of hip OA was rated equivalent to Kellgren-Lawrence grading system, however based on the ultrasound.

AP, anterior-posterior; KLG, Kellgren-Lawrence grading; OA, osteoarthritis.

grading was the operator's overall assessment of the findings mentioned above. Image examples of the different stages are illustrated in online supplemental file 2.

Lateral hip

Trochanter bursitis was defined as fluid in any bursa in the trochanter region.

When assessing the findings, if a rating was questionable, the finding was rated in the lowest category in question. Representative images were stored during the examinations. After 4–6 weeks, one of the operators (the chiropractor) reassessed the existing images, blinded to prior ratings and measurements in order to investigate intra-rater reliability. Only joint recess profile, BCD, overall joint effusion/synovitis, femoral osteophytes and femoral cartilage thickness were assessed a second time, since we found that the other findings (acetabular osteophytes, labral changes, femoral head deformation and

bursitis) required dynamic evaluation in order to be properly assessed.

X-rays

Anterior-posterior (AP) pelvic or hip (according to Department guidelines) images were recorded standing, unless the patient could not stand correctly (13 hips were recorded lying). An imaging specialist with 10 years of experience in musculoskeletal imaging blinded to the ultrasound findings assessed all the X-rays. The X-rays were scored for individual OA features in accordance with the Osteoarthritis Research Society International Atlas.²⁰ The grade of radiological hip OA was assessed using the KLG system.²¹ Radiographic hip OA was defined as a KLG ≥ 2 .

Statistical analysis

The statistical analysis was performed using STATA/IC V.15.1. Inter-rater and intra-rater reliability and agreement between ultrasound imaging and X-ray were assessed using Cohen's ordinal kappa statistics for binomial categorical variables and weighted kappa for ordered categorical variables. Quadratic weights were applied according to the number of categories and a 95% CI was calculated by bootstrap resampling with 1000 repetitions for ordered categorical variables. The interclass correlation coefficient (ICC) for agreement (absolute agreement, two-way random, single measures)²² was used to assess ratings on continuous scales. Bland-Altman plots with 95% limits of agreement (LOA) were calculated to evaluate systematic differences, with the 95% LOA calculated as the mean difference $\pm 1.96 \times \text{SD}$ of the difference.²³

In the interpretation of the kappa coefficient, the Landis and Koch standards for strength of agreement were used: poor ($\kappa < 0.0$), slight ($0.0 \leq \kappa < 0.2$), fair ($0.2 \leq \kappa < 0.4$), moderate ($0.4 \leq \kappa < 0.6$), substantial ($0.6 \leq \kappa < 0.8$) and almost perfect ($0.8 \leq \kappa \leq 1$).²⁴ The ICC for agreement was interpreted as follows: $\text{ICC} < 0.5 = \text{poor}$, $0.5 \leq \text{ICC} \leq 0.75 = \text{moderate}$, $0.75 < \text{ICC} \leq 0.9 = \text{good}$ and $> 0.9 = \text{excellent}$.²²

Patient and public involvement

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

Bilateral hips in 50 participants ($n=92$) were included in the study. Due to previous surgery, eight non-painful hips were excluded. Of the included participants, 43 were referred from general practitioners and 7 from orthopaedic surgeons, 32 (64%) were women and 26 (52%) had symptoms for more than 16 weeks. Age ranged from 42 to 90 years (median 67 years), mean body mass index was 26.9 (range 18.4–36.6). Mean HOOS on pain and function in daily living was 49 (SD 19) and 53 (SD 19), respectively (100=normal function). Because we followed the Department guidelines all participants had an X-ray

Table 3 Inter-rater reliability and agreement between two ultrasound operators on ultrasound findings in 92 hips

Ultrasound finding	Observed agreement (%)	Expected agreement (%)	Kappa (95% CI)	ICC (95% CI)
Joint recess profile	95	79	0.74 (0.58 to 0.85)	
Bone-capsule distance (including capsule)	98	87		0.88 (0.82 to 0.93)
Bone-capsule distance (excluding capsule)	98	88		0.89 (0.84 to 0.93)
Overall joint effusion/synovitis	97	68	0.90 (0.79 to 1.00)	
Femoral osteophytes	93	82	0.59 (0.45 to 0.70)	
Acetabular osteophytes	91	85	0.39 (0.20 to 0.54)	
Femoral head deformation	95	85	0.70 (0.54 to 0.82)	
Femoral cartilage changes	94	83	0.64 (0.51 to 0.75)	
Femoral cartilage thickness*	96	91		0.59 (0.37 to 0.73)
Labrum changes†	89	73	0.60 (0.43 to 0.73)	
Iliopsoas bursitis	92	93	-0.04 (-0.09 to 0.01)	
Trochanter bursitis	95	93	0.26 (-0.18 to 0.70)	
OA grading	96	80	0.80 (0.71 to 0.87)	

*n=70 hips; due to exclusion of 22 hips with pronounced irregular articular cartilage.

†n=79 hips; due to exclusion of 13 hips with labrum not rated.

ICC, interclass correlation coefficient; OA, osteoarthritis.

of the painful hip, but only some had bilateral hip (pelvic AP) resulting in 63 hip X-rays. Of these, 36 (57%) had KLG 2 or more. This is defined as radiological OA. On an individual level 28 of the 50 participants had radiographic OA in either one or both hips.

Prevalence of ultrasound and X-ray findings are shown in tables 1 and 2, respectively. The most prevalent ultrasound finding was labrum changes (53%–57% had moderate or severe changes) and least prevalent findings were effusion in iliopsoas or trochanter bursas (2%–6%).

The strongest inter-rater reliability was found for BCD (ICC=0.9) regardless whether it was measured to the inner or outer edge of the capsule, overall evaluation of hip effusion/synovitis ($\kappa=0.9$) and OA grading ($\kappa=0.8$). Acetabular and femoral osteophytes, femoral head deformation, femoral cartilage changes and labrum changes

had $\kappa=0.4$ – 0.7 (table 3). Trochanter bursitis had $\kappa=0.3$ and iliopsoas bursitis had $\kappa=0$.

Intra-rater reliability of interpretation of captured images had equal or higher values compared with inter-rater reliability (table 4).

The mean difference between the operators on numeric measures was 0.3 mm (95% CI 0.1 to 0.6) for BCD (outer edge of the joint capsule) and -0.1 mm (95% CI -0.17 to -0.05) for cartilage thickness. Bland-Altman plots for these measures (figure 2) showed a few outliers, but no funnel effects (increasing difference with increasing mean size).

Agreement between ultrasound and X-ray findings on femoral head deformation and grading of OA was $\kappa=0.5$

Table 4 Intrarater reliability and agreement on ultrasound findings* assessed with 4–6 weeks of interval by operator A

Ultrasound finding	Observed agreement (%)	Expected agreement (%)	Kappa (95% CI)	ICC (95% CI)
Joint recess profile	93	78	0.70 (0.47 to 0.86)	
Bone-capsule distance (including capsule)	100	88		0.99 (0.98 to 0.99)
Bone-capsule distance (excluding capsule)	100	88		0.99 (0.99 to 0.99)
Overall joint effusion/synovitis	96	67	0.87 (0.74 to 0.99)	
Femoral osteophytes	95	75	0.81 (0.72 to 0.88)	
Femoral cartilage thickness†	99	93		0.93 (0.88 to 0.95)

*Only the listed findings were assessed a second time, since we found that the other findings required dynamic evaluation in order to be properly assessed.

†n=65 hips, due to exclusion of 27 hips with pronounced irregular articular cartilage.

ICC, interclass correlation coefficient.

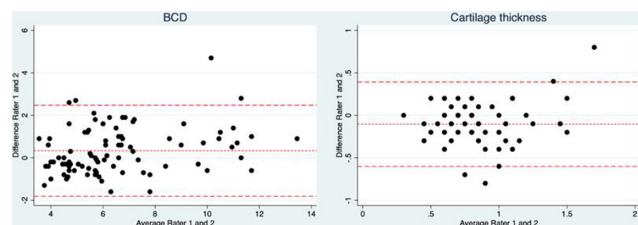


Figure 2 Bland-Altman plot with 95% limits of agreement for the two operators' recordings of bone-capsule distance (BCD) and cartilage thickness.

for both operators. For femoral and acetabular osteophytes, κ ranged from 0.2 to 0.4 (table 5).

DISCUSSION

Due to the dynamic nature of ultrasound, the difference in acquisition technique between operators is an important concern when using ultrasound examinations in both research and clinical settings. However, the few previous studies on reliability of ultrasound findings in patients with hip OA have only assessed reliability using recorded film and images. Thus, to our knowledge this is the first study to include differences in acquisition of images between the two operators in the assessment of reliability. We found substantial to almost perfect inter-rater reliability for findings related to effusion/synovitis and for the most common findings related to OA. In contrast, acetabular osteophytes had moderate, trochanter effusion had fair and iliopsoas bursitis had poor reliability. Overall, these results support ultrasound imaging as a reliable diagnostic tool in hip OA assessment.

Hip effusion/synovitis can be assessed in several ways: evaluation of BCD, evaluation of the joint recess profile or an overall evaluation. In this study, evaluation of BCD and an overall evaluation performed similarly, with excellent inter-rater and intrarater reliability for BCD and almost perfect for evaluation of effusion/synovitis overall, in line with another recent study.²⁵ However, evaluation of the joint recess profile had only substantial inter-rater and intrarater agreement. The prevalence of effusion/synovitis was almost identical regardless of which method we used for evaluation (20%–22%). These results support

the use of BCD as well as an overall evaluation when assessing hip joint effusion/synovitis.

Studies investigating diagnostic accuracy of ultrasound imaging in relation to effusion/synovitis and labral tears (using MRI and MR arthrography as the reference standard) report significant correlations between ultrasound and MRI on effusion/synovitis,²⁵ and different results in diagnosing labral tears.^{26 27} This, combined with excellent inter-rater and intrarater reliability for effusion/synovitis and substantial inter-rater reliability for labrum changes ($\kappa=0.6$), demonstrated in the current study support for ultrasound being an alternative to MRI when investigating effusion and synovitis in the hip and, with some precaution, labral changes. Osseous structures such as femoral osteophytes and femoral head abnormality have also been assessed previously, and the studies report moderate to substantial inter-rater reliability ($\kappa=0.4$ – 0.7) in line with our findings.^{16 28}

The prevalence and reliability of iliopsoas bursitis assessed with ultrasound has previously been evaluated in a retrospective study including 860 patients with symptomatic and radiological hip OA (KLG 2–4).⁷ The authors found a prevalence of iliopsoas bursitis of 2.2% and a perfect inter-rater reliability ($\kappa=1$). Using the same criteria for diagnosis of iliopsoas bursitis,¹⁷ we found a prevalence of 3%–4% but poor inter-rater reliability, probably because agreement on the presence of bursal effusion is easier on an existing image versus on real-time images. Furthermore, we found small iliopsoas effusions, which can be difficult to diagnose with ultrasound.

The trochanter region was only investigated for bursitis. The intention was solely to investigate for obvious differential diagnosis to lateral hip pain, since OA changes were our primary interest. In the planning of the study we considered several definitions for trochanteric bursitis, but the literature is sparse. A commonly used definition is whether there is fluid in the bursa and therefore only cases with bursal effusion were encountered as bursitis.²⁹ Undifferentiated rating may have influenced the prevalence of trochanter bursitis (2%–6%).

Intra-rater reliability had an equal or slightly higher reliability coefficient, compared with inter-rater reliability, which is to be expected because intrarater reliability is

Table 5 Agreement between ultrasound and radiographic findings (n=63)

Findings rated on ultrasound and X-rays	Ultrasound operator A			Ultrasound operator B		
	Observed agreement (%)	Expected agreement (%)	Kappa (95% CI)	Observed agreement (%)	Expected agreement (%)	Kappa (95% CI)
Femoral osteophytes	85	80	0.24 (0.08 to 0.41)	83	79	0.21 (0.04 to 0.40)
Acetabular osteophytes	90	84	0.37 (0.14 to 0.55)	90	86	0.26 (0.01 to 0.48)
Femoral head deformation	92	88	0.35 (0.08 to 0.58)	93	87	0.42 (0.17 to 0.66)
OA grading	90	79	0.51 (0.34 to 0.68)	89	80	0.46 (0.27 to 0.64)

OA, osteoarthritis.

usually higher and image acquisition was eliminated as a source of variation, since intrarater reliability was assessed on recorded images.

Agreement between single osseous findings assessed on ultrasound and X-ray was only fair to moderate for individual findings ($\kappa=0.2-0.5$) as expected when comparing two different modalities. In relation to the grade of OA in general, we found moderate agreement ($\kappa=0.5$). While ultrasound can visualise inflammatory and subtle changes in the anterior femoral cartilage, X-ray gives a better insight into osseous changes. Ultrasound and radiographs may not detect the same structural lesions and thus it would not necessarily be the same osteophytes the two modalities assess. However, further investigation is needed to determine differences in relation to association with symptoms and prognosis between OA grading on ultrasound and X-ray.

Strengths and limitations

Our study has some limitations. Intra-rater reliability did not include findings that required dynamic examination. The X-rays were recorded according to the Department guidelines, and therefore there was some variation in acquisition procedures. The difference in load distribution between standing and supine recordings might have affected the degree of joint space narrowing, and potentially other structural findings.

One of the strengths of the current study was including the acquisition technique in the assessment of reliability between observers, as each ultrasound operator independently examined and assessed each hip. Another strength is the application of the OMERACT validated ultrasound definitions for osteophytes, cartilage, effusion and synovial hypertrophy making the findings applicable for other clinicians. Moreover, the participants in this study are representative of a broad spectrum of patients with hip pain and suspected hip OA, making the results transferable to other clinical settings. However, both operators were experienced in term of clinical knowledge and scanning techniques and since hip joint ultrasound is considered to be challenging our findings may not apply to inexperienced clinicians.

CONCLUSION

This study demonstrated excellent inter-rater reliability of ultrasound findings related to hip effusion/synovitis and substantial to almost perfect inter-rater reliability on the most common ultrasound findings related to hip OA and OA grading. Agreement between OA grading rated on ultrasound and X-rays was moderate. Overall, these results support ultrasound imaging as a reliable tool in the assessment of hip OA.

Author affiliations

¹Department of Sports Science and Clinical Biomechanics, University of Southern Denmark Faculty of Health Sciences, Odense, Denmark

²Department of Radiology, Lillebaelt Hospital Vejle, Vejle, Denmark

³Diagnostic Centre, University Research Clinic for Innovative Patient Pathways, Silkeborg Regional Hospital, Silkeborg, Denmark

⁴The Rheumatology Research Unit, Odense University Hospital, Odense, Denmark

⁵Copenhagen Center for Arthritis Research and Center for Rheumatology and Spine Diseases, Rigshospitalet Glostrup, Glostrup, Denmark

⁶Department of Clinical Medicine, University of Copenhagen Faculty of Health and Medical Sciences, Copenhagen, Denmark

⁷Nordic Institute of Chiropractic and Clinical Biomechanics, Odense, Denmark

Acknowledgements The authors would like to acknowledge Lise Bolander Malvang, science radiographer, and Karin Kronborg Andersen, coordinating specialist in conventional radiography at the radiology department of Silkeborg Regional Hospital for invaluable help with recruitment and practical execution of the study. We also like to acknowledge Susanne Brogaard Krogh, musculoskeletal specialist, who assessed the X-rays, and Odense Patient data Explorative Network (OPEN), Odense University Hospital, Odense, Denmark.

Contributors SC, JH, BA, LT and UF conceived and designed the study protocol. SC and SK performed the ultrasound examinations. SC performed the second assessment of images for the purpose of intrarater agreement. SC, JH and BA planned the statistical analyses and SC performed the analysis. SC drafted the manuscript with substantial contribution from all authors. All authors read and approved the final manuscript.

Funding This work was supported by the Danish Chiropractic Research Foundation (16/3065), the Region of Southern Denmark (17/33620) and the IMK public fund (30206-353).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The study was conducted according to the Declaration of Helsinki and Danish legislation and before study inclusion, each patient gave written informed consent for research use and publication of their anonymised data. The Regional Scientific Ethics Committee for Southern Denmark determined that under Danish law, this study did not require formal ethics approval (project ID S-20180107).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. No additional data are available.

This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Stine Clausen <http://orcid.org/0000-0002-8632-4596>

REFERENCES

- Glyn-Jones S, Palmer AJR, Agricola R, *et al*. Osteoarthritis. *The Lancet* 2015;386:376–87.
- Huang BK, Tan W, Scherer KF, *et al*. Standard and advanced imaging of hip osteoarthritis. what the radiologist should know. *Semin Musculoskelet Radiol* 2019;23:289–303.
- Hunter DJ, Arden N, Conaghan PG, *et al*. Definition of osteoarthritis on MRI: results of a Delphi exercise. *Osteoarthritis Cartilage* 2011;19:963–9.
- Iagnocco A, Naredo E. Ultrasound of the osteoarthritic joint. *Clin Exp Rheumatol* 2017;35:527–34.

- 5 Sudula SN. Imaging the hip joint in osteoarthritis: a place for ultrasound? *Ultrasound* 2016;24:111–8.
- 6 Oo WM, Linklater JM, Daniel M, *et al.* Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. *Osteoarthritis Cartilage* 2018;26:601–11.
- 7 Tormenta S, Sconfienza LM, Iannessi F, *et al.* Prevalence study of iliopsoas bursitis in a cohort of 860 patients affected by symptomatic hip osteoarthritis. *Ultrasound Med Biol* 2012;38:1352–6.
- 8 Poulsen E, Overgaard S, Vestergaard JT, *et al.* Pain distribution in primary care patients with hip osteoarthritis. *Fam Pract* 2016;33:601–6.
- 9 Kottner J, Audigé L, Brorson S, *et al.* Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
- 10 Henrica CW, de Vet CBT, Mokkink LB, *et al.* *Measurement in medicine (practical guides to biostatistics and epidemiology.* Cambridge University Press, 2011.
- 11 Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster universities osteoarthritis index. *Scand J Rheumatol* 2003;32:46–51.
- 12 Möller I, Janta I, Backhaus M, *et al.* The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Ann Rheum Dis* 2017;76:1974–9.
- 13 The minimum training recommendations for the practice of medical ultrasound. *Eur J Ultrasound*;29:94–6.
- 14 Bruyn GA, Iagnocco A, Naredo E, *et al.* OMERACT definitions for ultrasonographic pathologies and elementary lesions of rheumatic disorders 15 years on. *J Rheumatol* 2019;46:1388–93.
- 15 Martinoli C, Bianchi S. *Ultrasound of the musculoskeletal system.* Springer-Verlag Berlin Heidelberg, 2007.
- 16 Qvistgaard E, Torp-Pedersen S, Christensen R, *et al.* Reproducibility and inter-reader agreement of a scoring system for ultrasound evaluation of hip osteoarthritis. *Ann Rheum Dis* 2006;65:1613–9.
- 17 Wunderbaldinger P, Bremer C, Schellenberger E, *et al.* Imaging features of iliopsoas bursitis. *Eur Radiol* 2002;12:409–15.
- 18 Koski JM, Anttila PJ, Isomäki HA. Ultrasonography of the adult hip joint. *Scand J Rheumatol* 1989;18:113–7.
- 19 Machado FS, Natour J, Takahashi RD, *et al.* Articular Ultrasound in Asymptomatic Volunteers: Identification of the Worst Measures of Synovial Hypertrophy, Synovial Blood Flow and Joint Damage Among Small-, Medium- and Large-Sized Joints. *Ultrasound Med Biol* 2017;43:1141–52.
- 20 Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage* 2007;15 Suppl A:A1–56.
- 21 Kellgren JH, Lawrence JS. Radiological assessment of osteoarthritis. *Ann Rheum Dis* 1957;16:494–502.
- 22 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15:155–63.
- 23 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
- 24 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- 25 Steer KJD, Bostick GP, Woodhouse LJ, *et al.* Can effusion-synovitis measured on ultrasound or MRI predict response to intra-articular steroid injection in hip osteoarthritis? *Skeletal Radiol* 2019;48:227–37.
- 26 Jin W, Kim KI, Rhyu KH, *et al.* Sonographic evaluation of anterosuperior hip labral tears with magnetic resonance arthrographic and surgical correlation. *J Ultrasound Med* 2012;31:439–47.
- 27 Troelsen A, Mechlenburg I, Gelineck J, *et al.* What is the role of clinical tests and ultrasound in acetabular labral tear diagnostics? *Acta Orthop* 2009;80:314–8.
- 28 Abraham AM, Pearce MS, Mann KD, *et al.* Population prevalence of ultrasound features of osteoarthritis in the hand, knee and hip at age 63 years: the Newcastle thousand families birth cohort. *BMC Musculoskelet Disord* 2014;15:162.
- 29 Long SS, Surrey DE, Nazarian LN. Sonography of greater trochanteric pain syndrome and the rarity of primary bursitis. *AJR Am J Roentgenol* 2013;201:1083–6.