



BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis. An inter- and intra-rater reliability study.

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-038643
Article Type:	Original research
Date Submitted by the Author:	19-Mar-2020
Complete List of Authors:	Clausen, Stine; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Lillebaelt Hospital Vejle, Department of Radiology Kjær, Søren; Silkeborg Regional Hospital, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways Fredberg, Ulrich; Silkeborg Regional Hospital, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways; Odense University Hospital, The Rheumatology Research Unit Terslev, Lene; Rigshospitalet Glostrup, Copenhagen Center for Arthritis Research and Center for Rheumatology and Spine Diseases; University of Copenhagen Faculty of Health and Medical Sciences, Department of Clinical Medicine Hartvigsen, Jan; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Nordic Institute of Chiropractic and Clinical Biomechanics Arnbak, Bodil; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Lillebaelt Hospital Vejle, Department of Radiology
Keywords:	Ultrasound < RADIOLOGY & IMAGING, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Diagnostic radiology < RADIOLOGY & IMAGING, RHEUMATOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis. An inter- and intra-rater reliability study.

Stine Clausen^{1,2}
Søren Geil Kjær³
Ulrich Fredberg^{3,4}
Lene Terslev^{5,6}
Jan Hartvigsen^{1,7}
Bodil Arnbak^{1,2}

Corresponding author: Stine Clausen, University of Southern Denmark, Campusvej 55, 5230 Odense M, sclausen@health.sdu.dk telephone +45 30256613 Fax: none

¹ Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

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

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

⁴ The Rheumatology Research Unit, Odense University Hospital, University of Southern Denmark, Denmark.

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

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

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

Keywords: Ultrasound, Hip, Osteoarthritis, Reliability

Word count, excluding title page, abstract, references, figures and tables: 2696

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objectives: The objectives of this study were to assess 1) inter- and intra-rater reliability of ultrasound imaging in patients with hip osteoarthritis, and 2) agreement between ultrasound and x-ray findings of hip osteoarthritis using validated OMERACT ultrasound definitions for pathology.

Design: An inter-rater and intra-rater 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- and intra-rater reliability and agreement between ultrasound imaging and x-ray were assessed using Cohen's ordinal kappa statistics for binominal categorical variables and weighted kappa for ordered categorical variables.

Results: Kappa values (κ) for inter-rater reliability was 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 0.4-0.7. Intra-rater reliability had κ equal or higher compared to inter-rater reliability. Agreement between ultrasound and x-ray findings ranged from κ 0.2-0.5.

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.

Strengths and limitations of this study

- Using OMERACT 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 OA, making the results transferable to other clinical settings.
- Intra-rater 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's guidelines, and variation in acquisition procedures might have affected the findings.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Osteoarthritis (OA) is characterised by progressive destruction of articular cartilage, changes in bone tissue, osteophyte formation 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 to visualise both structural changes in the articular bone and inflammatory changes in soft tissue around the joint. MRI has the advantage of revealing intra-articular structures better than all other modalities; however, access to MRI can be restricted due to its high expense and limited availability. Ultrasound, on the other hand, is inexpensive and relatively accessible and allows for dynamic examination, assessment of Doppler activity and clinician-patient interaction during examination. 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[2, 3]. Iliopsoas bursitis rarely occurs, but prevalence increases at more advanced stages of OA[4]. Acceptable reliability of ultrasound-specific lesions provides the foundation for diagnostic or epidemiologic 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[3]. 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 is needed.

The primary objective of this study was to assess the inter- and intra-rater reliability of ultrasound findings in patients with hip OA. The trochanter region was also investigated, as patients with hip pain often complain of pain in this location[5]. The secondary aim was to assess agreement between ultrasound and radiologic findings related to hip OA.

MATERIALS AND METHODS

Guidelines for Reporting Reliability and Agreement Studies[6] 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[7].

All participants completed an electronic questionnaire containing demographic data, and the Danish version of the Hip disability and Osteoarthritis Outcome Score (HOOS) questionnaire[8], 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 a 18-5 MHz linear transducer. Predefined settings were used, with individual adjustment of the grey-scale settings and focus. The examinations were performed continuously, based on a protocol defined by The European League Against Rheumatism[9]. 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. Total study time (both hips, two raters) was approximately one hour.

The ultrasound operators were a chiropractor and a rheumatologist. Both had 10-15 year experience using musculoskeletal ultrasound (ultrasound qualification equivalent to European Federation of Societies for Ultrasound in Medicine and Biology level 2)[10]. 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[11]. Labrum changes and femoral head deformation were scored semiquantitatively (none, mild, moderate or severe) and trochanter and iliopsoas bursitis were scored dichotomously (present/absent). The ultrasound findings assessed, grading systems and definitions are listed below and in Supplementary file 1.

Ultrasound examination 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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

was made as close 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 (Fig. 1). We measured BCD from the cortical surface of the femoral neck to both the inner and the outer edge of the joint capsule, the latter combining joint fluid and synovium/capsule. BCD increase to 7 mm or more (inner edge of joint capsule) or a bilateral difference of 1 mm indicates effusion according to Koski criteria[12]. 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[13]. 3) An overall assessment of effusion/synovitis was performed based on the joint recess profile and BCD using Koski criteria[12], and the possible presence of hypoechoic or anechoic fluid in the joint recess along the femoral neck, and recorded as present or absent.

Subchondral bone cysts were assessed and rated according to a scoring system for the shape of the femoral head described by Qvistgaard et al.[14]. Changes in labrum were assessed according to Martinoli et al.[15]. Iliopsoas bursitis was diagnosed using Wunderbaldinger criteria[16].

At the end of the examination, the operator rated the degree of hip OA 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.

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

Where a rating was questionable, the finding was rated in the lowest category in question. Representative images were stored during the examinations. After 4 to 6 weeks, one of the operators reassessed the existing images, blinded to prior ratings and measurements in order to investigate intra-rater reliability. Findings that required the dynamic nature of ultrasound were not assessed a second time.

X-rays

Anterior-posterior (AP) pelvic or hip images were recorded standing, unless the patient could not stand correctly (13 hips were recorded lying). An imaging specialist 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[17]. The grade of radiologic hip OA was assessed using Kellgren-Lawrence grading system (KLG)[18]. Radiographic hip osteoarthritis was defined as a KLG ≥ 2 .

Statistical analysis

The statistical analysis was performed using STATA/IC 15.1. Inter- and intra-rater reliability and agreement between ultrasound imaging and x-ray were assessed using Cohen's ordinal kappa statistics for binominal categorical variables and weighted kappa for ordered categorical variables. Quadratic weights were applied according to the number of categories and a 95% confidence interval (CI) was calculated by bootstrap resampling with 1000 repetitions for ordered categorical variables. The interclass correlation coefficient for agreement (ICC, absolute agreement, two-way random, single measures)[19] was used to assess ratings on continuous scales and Bland-Altman plots with 95% limits of agreement were used to evaluate systematic differences between the two operators as well as equal distribution of difference across the scale range[20].

In the interpretation of the kappa coefficient, the Landis and Koch standards for strength of agreement were used: poor ($\kappa < 0.0$), slight ($0.0 \leq \kappa \leq 0.2$), fair ($0.2 < \kappa \leq 0.4$), moderate ($0.4 < \kappa \leq 0.6$), substantial ($0.6 < \kappa \leq 0.8$) and almost perfect ($0.8 < \kappa \leq 1$)[21]. The ICC for agreement was interpreted as follows: $ICC < 0.5$ = poor, $0.5 \leq ICC \leq 0.75$ = moderate, $0.75 < ICC \leq 0.9$ = good and > 0.9 = excellent[19].

Ethics

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).

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 BMI was 26.9 (range 18.4 - 36.6). Mean HOOS on pain and function in daily living were 49 (SD 19) and 53 (SD 19), respectively (100=normal function). Because we followed Department guidelines all participants

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

had an x-ray 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.

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%).

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:		
<i>concave</i>	71 (77%)	73 (79%)
<i>straight</i>	12 (13%)	10 (11%)
<i>convex</i>	9 (10%)	9 (10%)
<i>not accessible</i>	0	0
Overall joint effusion/synovitis:		
<i>absent</i>	74 (80%)	73 (79%)
<i>present</i>	18 (20%)	19 (21%)
<i>not accessible</i>	0	0
Femoral osteophytes:		
<i>none</i>	23 (25%)	18 (20%)
<i>mild</i>	41 (45%)	43 (47%)
<i>moderate</i>	19 (21%)	22 (24%)
<i>severe</i>	9 (10%)	9 (10%)
<i>not accessible</i>	0	0
Acetabular osteophytes:		
<i>none</i>	38 (41%)	40 (44%)
<i>mild</i>	31 (34%)	38 (42%)
<i>moderate</i>	18 (20%)	11 (12%)
<i>severe</i>	5 (5%)	1 (1%)
<i>not accessible</i>	0	2 (1%)
Femoral head deformation:		
<i>none</i>	57 (62%)	49 (53%)
<i>mild</i>	23 (25%)	29 (32%)
<i>moderate</i>	8 (9%)	10 (11%)
<i>severe</i>	4 (4%)	4 (4%)
<i>not accessible</i>	0	0
Femoral cartilage changes:		
<i>none</i>	33 (36%)	25 (27%)
<i>mild</i>	31 (34%)	36 (39%)
<i>moderate</i>	23 (25%)	27 (29%)
<i>severe</i>	5 (5%)	4 (4%)
<i>not accessible</i>	0	0
Labrum changes:		

<i>none</i>	17 (18%)	9 (10%)
<i>mild</i>	21 (23%)	18 (20%)
<i>moderate</i>	17 (18%)	19 (21%)
<i>severe</i>	32 (35%)	33 (36%)
<i>not accessible</i>	5* (5%)	13* (14%)
Iliopsoas bursitis:		
<i>absent</i>	89 (97%)	88 (96%)
<i>present</i>	3 (3%)	4 (4%)
<i>not accessible</i>	0	0
Trochanter bursitis:		
<i>absent</i>	90 (98%)	86 (94%)
<i>present</i>	2 (2%)	5 (6%)
<i>not accessible</i>	0	0
OA grading:		
<i>none</i>	32 (35%)	27 (29%)
<i>mild</i>	27 (29%)	35 (38%)
<i>moderate</i>	25 (27%)	22 (24%)
<i>severe</i>	8 (9%)	8 (9%)
<i>not accessible</i>	0	0
	Mean distance Operator A	Mean distance Operator B
Bone-capsule distance (incl. capsule) (mm):	6.7 (SD 2.5)	6.4 (SD 2.2)
Bone-capsule distance (excl. 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.

Table 2. Prevalence of osseous findings and the grade of OA in the 63 hips that were x-rayed.

Findings rated on ultrasound and x-rays	Prevalence Operator A	Prevalence Operator B	Prevalence Radiographs
Femoral osteophytes:			
<i>none</i>	16 (25%)	14 (22%)	44 (70%)
<i>mild</i>	27 (43%)	28 (44%)	16 (25%)
<i>moderate</i>	15 (24%)	15 (24%)	2 (3%)
<i>severe</i>	5 (8%)	6 (10%)	1 (2%)
Acetabular osteophytes:			
<i>none</i>	27 (43%)	26 (42%)	19 (30%)
<i>mild</i>	21 (33%)	28 (45%)	29 (46%)
<i>moderate</i>	12 (19%)	7 (11%)	13 (21%)
<i>severe</i>	3 (5%)	1 (2%)	2 (3%)
Femoral head deformation:			
<i>none</i>	39 (62%)	33 (52%)	40 (63%)
<i>mild</i>	17 (27%)	21 (33%)	18 (29%)
<i>moderate</i>	5 (8%)	7 (11%)	2 (6%)
<i>severe</i>	2 (3%)	2 (3%)	1 (2%)
OA grading:			

<i>none; KLG 0&1</i>	23 (37%)	20 (32%)	27 (43%)
<i>mild; KLG 2</i>	18 (29%)	21 (33%)	17 (27%)
<i>moderate; KLG 3</i>	17 (27%)	17 (27%)	15 (24%)
<i>severe; KLG 4</i>	5 (8%)	5 (8%)	4 (6%)

KLG; Kellgren-Lawrence Grade.

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 κ 0.4 - 0.7 (Table 3). Trochanter bursitis had $\kappa=0.3$ and iliopsoas bursitis $\kappa=0$.

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

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 (0.58-0.85)	
Bone-capsule distance (incl. capsule)	98%	87%		0.88 (0.82-0.93)
Bone-capsule distance (excl. capsule)	98%	88%		0.89 (0.84-0.93)
Overall joint effusion/synovitis	97 %	68%	0.90 (0.79-1.00)	
Femoral osteophytes	93%	82%	0.59 (0.45-0.70)	
Acetabular osteophytes	91%	85%	0.39 (0.20- 0.54)	
Femoral head deformation	95%	85%	0.70 (0.54- 0.82)	
Femoral cartilage changes	94%	83%	0.64 (0.51- 0.75)	
Femoral cartilage thickness*	96%	91%		0.59 (0.37-0.73)
Labrum changes**	89%	73%	0.60 (0.43-0.73)	

Iliopsoas bursitis	92%	93%	-0.04 (-0.09-0.01)	
Trochanter bursitis	95%	93%	0.26 (-0.18-0.70)	
OA grading	96%	80%	0.80 (0.71-0.87)	

* 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.

Table 4. Intra-rater reliability and agreement on ultrasound findings assessed with a 4-6 weeks 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 – 0.86)	
Bone-capsule distance (incl. capsule)	100%	88%		0.99 (0.98 – 0.99)
Bone-capsule distance (excl. capsule)	100%	88%		0.99 (0.99 – 0.99)
Overall joint effusion/synovitis	96%	67%	0.87 (0.74 – 0.99)	
Femoral osteophytes	95%	75%	0.81 (0.72 - 0.88)	
Femoral cartilage thickness*	99%	93%		0.93 (0.88 – 0.95)

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

Bland-Altman plots for the numeric measures (Fig. 2) showed no systematic differences between the two operators and equal difference across the range.

Agreement between ultrasound and x-ray findings on femoral head deformation and grading of OA were $\kappa=0.5$ for both operators. For femoral and acetabular osteophytes, κ ranged from 0.2 to 0.4 (Table 5).

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 – 0.41)	83%	79%	0.21 (0.04 – 0.40)

Acetabular osteophytes	90%	84%	0.37 (0.14 – 0.55)	90%	86%	0.26 (0.01 – 0.48)
Femoral head deformation	92%	88%	0.35 (0.08 – 0.58)	93%	87%	0.42 (0.17 – 0.66)
OA grading	90%	79%	0.51 (0.34 – 0.68)	89%	80%	0.46 (0.27 – 0.64)

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 similar, with excellent inter- and intra-rater reliability for BCD and almost perfect for evaluation of effusion/synovitis overall, in line with another recent study[22]. However, evaluation of the joint recess profile had only substantial inter- and intra-rater 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[22], and different results in diagnosing labral tears[23, 24]. This combined with excellent inter- and intra-rater 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 has also been assessed previously, and the studies report moderate to substantial inter-rater reliability ($\kappa=0.4-0.7$) in line with our findings[14, 25].

The prevalence and reliability of iliopsoas bursitis assessed with ultrasound has previously been evaluated in a retrospective study including 860 patients with symptomatic and radiologic hip OA (KLG 2-4)[4]. 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[16], we found 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.

Intra-rater reliability had an equal or slightly higher reliability coefficient, compared with inter-rater reliability, which is to be expected because intra-rater reliability is usually higher and image acquisition was eliminated as a source of variation, since intra-rater 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 (κ 0.2-0.5) as expected when comparing two different modalities. While ultrasound is better at identifying small femoral osteophytes, x-ray is more likely to detect more profound osseous changes, thus it would not necessarily be the same osteophytes the two modalities assess. In relation to the grade of OA in general, we found a slightly better agreement (κ 0.5). This may be explained by each modality assess the level osteoarthritis, even though they use different types of findings in the assessment. While ultrasound can also visualize inflammatory changes and subtle changes in the anterior femoral cartilage, x-ray only visualize osseous changes. However, further investigation is needed to determine differences in relation to association with symptoms and prognosis between OA grading on UL 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 Department's guidelines, and therefore there was some variation in acquisition procedures. The difference in load distribution between standing and

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

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 Contributions SC, JH, BA and UF conceived and designed the study protocol. SC and SK performed the ultrasound examinations. SC and BA performed the statistical analyses. SC drafted the manuscript with substantial contribution from all authors. All authors read and approved the final manuscript.

Acknowledgements The authors would like to acknowledge Lise Bolander Malvang, science radiographer, Karin Kronborg Andersen, coordinating specialist in conventional radiography at the radiology department at 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 OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark.

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.

Data sharing statement No additional data are available.

Legends:

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

Fig. 2 Bland-Altman plot of the two operator's recordings of bone-capsule distance (BCD) and cartilage thickness.

REFERENCES

1. Glyn-Jones S, Palmer AJR, Agricola R et al. Osteoarthritis. *The Lancet* 2015; 386: 376-387. doi:[https://doi.org/10.1016/S0140-6736\(14\)60802-3](https://doi.org/10.1016/S0140-6736(14)60802-3)
2. Sudula SN. Imaging the hip joint in osteoarthritis: A place for ultrasound? *Ultrasound* 2016; 24: 111-118. doi:10.1177/1742271X16643118
3. Oo WM, Linklater JM, Daniel M et al. Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. *Osteoarthritis Cartilage* 2018. doi:10.1016/j.joca.2018.01.021. doi:10.1016/j.joca.2018.01.021
4. 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: 1352-1356. doi:10.1016/j.ultrasmedbio.2012.04.006
5. Poulsen E, Overgaard S, Vestergaard JT et al. Pain distribution in primary care patients with hip osteoarthritis. *Fam Pract* 2016; 33: 601-606. doi:10.1093/fampra/cmw071
6. Kottner J, Audige L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64: 96-106. doi:10.1016/j.jclinepi.2010.03.002
7. Henrica C. W. de Vet CBT, Lidwine B. Mokkink and Dirk L. Knol. *Measurement in medicine (Practical guides to Biostatistics And Epidemiology)*: Cambridge University Press; 2011
8. Klassbo 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
9. Moller I, Janta I, Backhaus M et al. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Ann Rheum Dis* 2017; 76: 1974-1979. doi:10.1136/annrheumdis-2017-211585
10. [Anonym]. The Minimum Training Recommendations for the Practice of Medical Ultrasound. *Eur J Ultrasound*; 29: 94-96
11. 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. doi:10.3899/jrheum.181095. doi:10.3899/jrheum.181095
12. Koski JM, Anttila PJ, Isomaki HA. Ultrasonography of the adult hip joint. *Scand J Rheumatol* 1989; 18: 113-117
13. 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: 1141-1152. doi:10.1016/j.ultrasmedbio.2017.01.021

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

14. 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: 1613-1619. doi:10.1136/ard.2005.050690

15. Carlo Martinoli SB. *Ultrasound of the Musculoskeletal System*: Springer-Verlag Berlin Heidelberg; 2007. doi:10.1007/978-3-540-28163-4

16. Wunderbaldinger P, Bremer C, Schellenberger E et al. Imaging features of iliopsoas bursitis. *Eur Radiol* 2002; 12: 409-415. doi:10.1007/s003300101041

17. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage* 2007; 15 Suppl A: A1-56. doi:10.1016/j.joca.2006.11.009

18. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis* 1957; 16: 494-502

19. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016; 15: 155-163. doi:10.1016/j.jcm.2016.02.012

20. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310

21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-174

22. 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: 227-237. doi:10.1007/s00256-018-3010-9

23. 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: 439-447

24. 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: 314-318. doi:10.3109/17453670902988402

25. 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. doi:10.1186/1471-2474-15-162

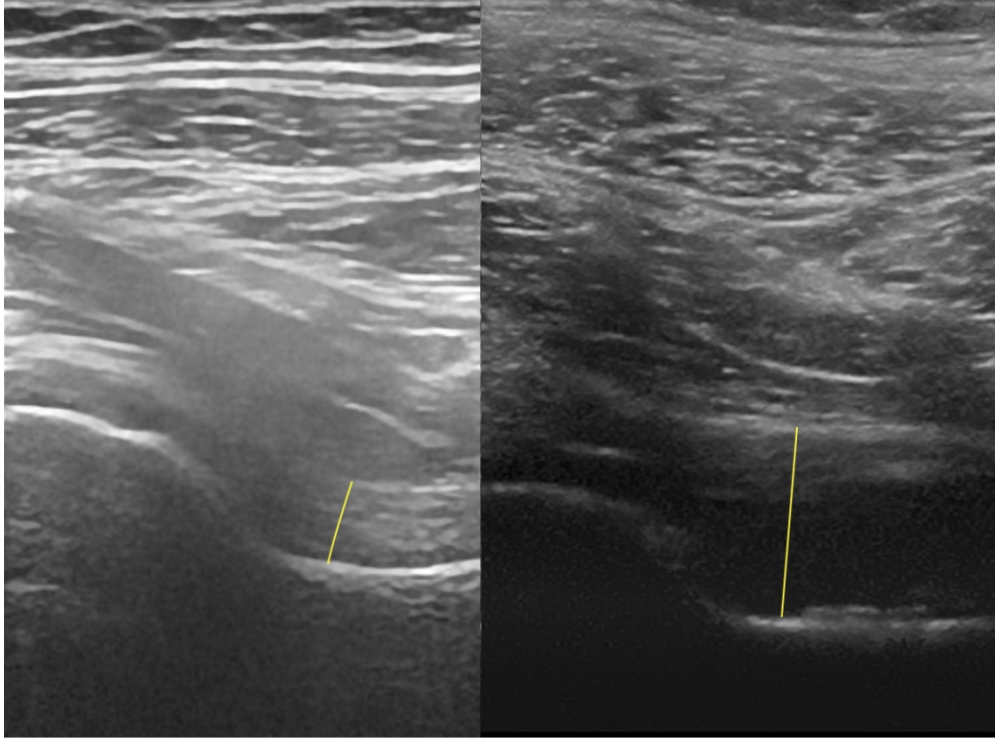


Figure 1 Anterior hip joint recess, longitudinal scan. The yellow line marks the bone-capsule distance. The hip on the right has effusion/synovitis.

100x74mm (300 x 300 DPI)

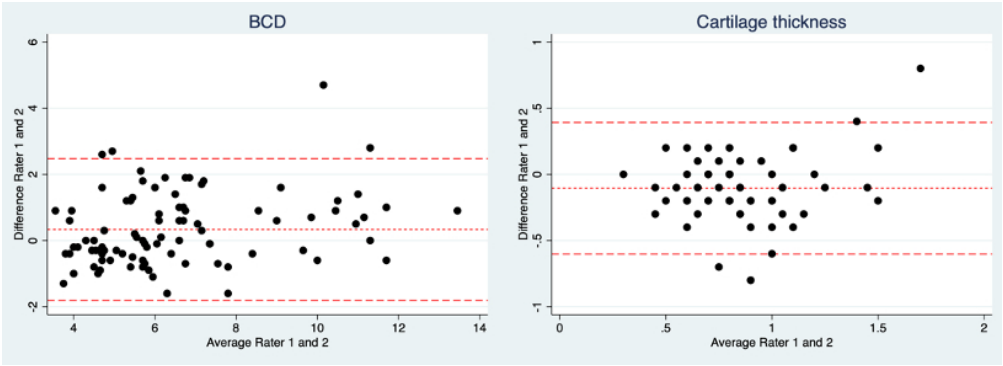


Figure 2 Bland-Altman plots with 95% limits of agreement for the two raters measurements of bone-capsule distance (BCD) and cartilage thickness.

68x24mm (300 x 300 DPI)

Supplementary file 1. Ultrasound findings in the hip, scoring systems and definitions used in the assessment.

Ultrasound finding	Scoring system <i>Binary, ordinal (semi-quantitatively) or numerical</i>	Definition
Joint recess profile	<i>(ordinal)</i> Concave Straight Convex	The course of the anterior joint recess along the anterior surface of the femoral neck
Bone-capsule distance (BCD) in the anterior joint recess	<i>(numerical)</i> Number in mm.	Distance from the outer femoral cortex to the outer edge of the capsule. A second measure was made from the femoral cortex to the inside edge of the capsule
Overall joint effusion/synovitis	<i>(binary)</i> Present / absent	Overall assessment of all the ultrasound findings related to effusion / synovitis
Femoral osteophytes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Acetabular osteophytes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Femoral head deformation	<i>(ordinal)</i> None, mild, moderate or severe	Normal (round), mild (slightly flattened), moderate (very flattened), severe (no obvious contour or the femoral head can be defined)
Femoral cartilage changes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Femoral cartilage thickness	<i>(numerical)</i> Number in mm.	Anterior surface of femoral head, as close to labrum as possible.
Labrum changes	<i>(ordinal)</i> None, mild, moderate or severe	Normal (homogeneous echogenicity), mild (heterogeneous echogenicity and labrum poorly defined), moderate (definite pathology) and severe (pathology or degeneration to a degree where labrum was not defined)
Iliopsoas bursitis	<i>(binary)</i> Present / absent	Fluid associated with the iliopsoas tendon ventral to the hip joint
Trochanter bursitis	<i>(ordinal)</i> Present / absent	Fluid in any bursa in the trochanter region

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

OA grading	(ordinal) None, mild, moderate or severe	The degree of hip OA based on the ultrasound findings
------------	--	--

Kottner J, Audige L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64: 96-106.

Guidelines for Reporting Reliability and Agreement Studies (GRRAS).		Page
TITLE AND ABSTRACT	1. Identify in title or abstract that inter-rater/intra-rater reliability or agreement was investigated.	1
INTRODUCTION	2. Name and describe the diagnostic or measurement device of interest explicitly.	4
	3. Specify the subject population of interest.	4
	4. Specify the rater population of interest (if applicable).	NA
	5. Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	4
METHODS	6. Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	5
	7. Describe the sampling method.	5
	8. Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	5 6
	9. State whether measurements/ratings were conducted independently.	5
	10. Describe the statistical analysis.	7
RESULTS	11. State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	5 7
	12. Describe the sample characteristics of raters and subjects (e.g. training, experience).	5 7-9
	13. Report estimates of reliability and agreement including measures of statistical uncertainty.	10
DISCUSSION	14. Discuss the practical relevance of results.	12
AUXILIARY MATERIAL	15. Provide detailed results if possible (e.g. online)	NA

BMJ Open

Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis. An inter- and intra-rater reliability study.

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-038643.R1
Article Type:	Original research
Date Submitted by the Author:	02-Jul-2020
Complete List of Authors:	Clausen, Stine; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Lillebaelt Hospital Vejle, Department of Radiology Kjær, Søren; Silkeborg Regional Hospital, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways Fredberg, Ulrich; Silkeborg Regional Hospital, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways; Odense University Hospital, The Rheumatology Research Unit Terslev, Lene; Rigshospitalet Glostrup, Copenhagen Center for Arthritis Research and Center for Rheumatology and Spine Diseases; University of Copenhagen Faculty of Health and Medical Sciences, Department of Clinical Medicine Hartvigsen, Jan; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Nordic Institute of Chiropractic and Clinical Biomechanics Arnbak, Bodil; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Lillebaelt Hospital Vejle, Department of Radiology
Primary Subject Heading:	Radiology and imaging
Secondary Subject Heading:	Rheumatology
Keywords:	Ultrasound < RADIOLOGY & IMAGING, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Diagnostic radiology < RADIOLOGY & IMAGING, RHEUMATOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis. An inter- and intra-rater reliability study.

Stine Clausen^{1,2}
Søren Geil Kjær³
Ulrich Fredberg^{3,4}
Lene Terslev^{5,6}
Jan Hartvigsen^{1,7}
Bodil Arnbak^{1,2}

Corresponding author: Stine Clausen, University of Southern Denmark, Campusvej 55, 5230 Odense M, sclausen@health.sdu.dk telephone +45 30256613 Fax: none

¹ Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

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

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

⁴The Rheumatology Research Unit, Odense University Hospital, University of Southern Denmark, Denmark.

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

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

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

Keywords: Ultrasound, Hip, Osteoarthritis, Reliability

Word count, excluding title page, abstract, references, figures and tables: 2947

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objectives: The objectives of this study were to assess 1) inter- and intra-rater reliability of ultrasound imaging in patients with hip osteoarthritis, and 2) agreement between ultrasound and x-ray findings of hip osteoarthritis using validated OMERACT ultrasound definitions for pathology.

Design: An inter-rater and intra-rater 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- and intra-rater reliability and agreement between ultrasound imaging and x-ray were assessed using Cohen's ordinal kappa statistics for binominal categorical variables and weighted kappa for ordered categorical variables.

Results: Kappa values (κ) for inter-rater reliability was 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 0.4-0.7. Intra-rater reliability had κ equal or higher compared to inter-rater reliability. Agreement between ultrasound and x-ray findings ranged from κ 0.2-0.5.

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.

Strengths and limitations of this study

- Using OMERACT 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 OA, making the results transferable to other clinical settings.
- Intra-rater 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's guidelines, and variation in acquisition procedures might have affected the findings.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Osteoarthritis (OA) is characterized by progressive destruction of articular cartilage, changes in bone tissue, osteophyte formation and joint inflammation resulting in loss of normal joint function[1]. The pathophysiological changes can be visualized 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 to visualize 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 modalities[3]; 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 epidemiologic 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 is needed.

The primary objective of this study was to assess the inter- 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 radiologic findings related to hip OA.

MATERIALS AND METHODS

Guidelines for Reporting Reliability and Agreement Studies[9] 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 a 18-5 MHz linear transducer (central frequency of 18 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 minutes for each examiner.

The ultrasound operators were a chiropractor and a rheumatologist. Both had 10-15 year 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 Reumatology (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 labrum was not defined). Femoral head deformation were assessed and rated semi-quantitatively (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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

whether there was effusion in the bursa (present/absent)[17]. The ultrasound findings assessed, grading systems and definitions are listed below and described in Supplementary file 1. Image examples are illustrated in Supplementary file 2.

Ultrasound examination 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 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 (Fig. 1). We measured BCD from the cortical surface of the femoral neck to both the inner and the outer edge of the joint capsule, the latter combining joint fluid and synovium/capsule. BCD increase to 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 hypoechoic 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 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 grading was the operator's overall assessment of the findings mentioned above. Image examples of the different stages are illustrated in Supplementary file 2.

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

Where a rating was questionable, the finding was rated in the lowest category in question. Representative images were stored during the examinations. After 4 to 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, bone-capsule distance, 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's guidelines) images were recorded standing, unless the patient could not stand correctly (13 hips were recorded lying). An imaging specialist with ten years' 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 radiologic hip OA was assessed using Kellgren-Lawrence grading system (KLG)[21]. Radiographic hip osteoarthritis was defined as a KLG ≥ 2 .

Statistical analysis

The statistical analysis was performed using STATA/IC 15.1. Inter- and intra-rater reliability and agreement between ultrasound imaging and x-ray were assessed using Cohen's ordinal kappa statistics for binominal categorical variables and weighted kappa for ordered categorical variables. Quadratic weights were applied according to the number of categories and a 95% confidence interval (CI) was calculated by bootstrap resampling with 1000 repetitions for ordered categorical variables. The interclass correlation coefficient for agreement (ICC, 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 ($0.0 \leq \kappa \leq 0.2$), fair ($0.2 < \kappa \leq 0.4$), moderate ($0.4 < \kappa \leq 0.6$), substantial ($0.6 < \kappa \leq 0.8$) and almost perfect ($0.8 < \kappa \leq 1$)[24]. The ICC for agreement was interpreted as follows: ICC < 0.5 = poor, $0.5 \leq \text{ICC} \leq 0.75$ = moderate, $0.75 < \text{ICC} \leq 0.9$ = good and > 0.9 = excellent[22].

Ethics

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).

Patient and Public Involvement

1
2
3
4 Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans
5 of our research.
6
7

8
9 **RESULTS**

10
11 Bilateral hips in 50 participants (n=92) were included in the study. Due to previous surgery, eight
12 non-painful hips were excluded. Of the included participants, 43 were referred from general
13 practitioners and 7 from orthopaedic surgeons, 32 (64%) were women and 26 (52%) had symptoms
14 for more than 16 weeks. Age ranged from 42 to 90 years (median 67 years), mean BMI was 26.9
15 (range 18.4 - 36.6). Mean HOOS on pain and function in daily living were 49 (SD 19) and 53 (SD
16 19), respectively (100=normal function). Because we followed Department guidelines all participants
17 had an x-ray of the painful hip, but only some had bilateral hip (pelvic AP) resulting in 63 hip x-rays.
18 Of these, 36 (57%) had KLG 2 or more. This is defined as radiological OA. On an individual level
19 28 of the 50 participants had radiographic OA in either one or both hips.
20
21 Prevalence of ultrasound and x-ray findings are shown in Tables 1 and 2, respectively. The most
22 prevalent ultrasound finding was labrum changes (53-57% had moderate or severe changes) and least
23 prevalent findings were effusion in iliopsoas or trochanter bursas (2-6%).
24
25
26
27
28
29
30
31
32

33
34 Table 1. Prevalence and mean measure of ultrasound findings in 50 participants (92 hips) examined
35 by two operators blinded to each other's findings.
36

Ultrasound finding	Prevalence Operator A	Prevalence Operator B
Joint recess profile:		
<i>concave</i>	71 (77%)	73 (79%)
<i>straight</i>	12 (13%)	10 (11%)
<i>convex</i>	9 (10%)	9 (10%)
<i>not accessible</i>	0	0
Overall joint effusion/synovitis:		
<i>absent</i>	74 (80%)	73 (79%)
<i>present</i>	18 (20%)	19 (21%)
<i>not accessible</i>	0	0
Femoral osteophytes:		
<i>none</i>	23 (25%)	18 (20%)
<i>mild</i>	41 (45%)	43 (47%)
<i>moderate</i>	19 (21%)	22 (24%)
<i>severe</i>	9 (10%)	9 (10%)
<i>not accessible</i>	0	0
Acetabular osteophytes:		
<i>none</i>	38 (41%)	40 (44%)

<i>mild</i>	31 (34%)	38 (42%)
<i>moderate</i>	18 (20%)	11 (12%)
<i>severe</i>	5 (5%)	1 (1%)
<i>not accessible</i>	0	2 (1%)
Femoral head deformation:		
<i>none</i>	57 (62%)	49 (53%)
<i>mild</i>	23 (25%)	29 (32%)
<i>moderate</i>	8 (9%)	10 (11%)
<i>severe</i>	4 (4%)	4 (4%)
<i>not accessible</i>	0	0
Femoral cartilage changes:		
<i>none</i>	33 (36%)	25 (27%)
<i>mild</i>	31 (34%)	36 (39%)
<i>moderate</i>	23 (25%)	27 (29%)
<i>severe</i>	5 (5%)	4 (4%)
<i>not accessible</i>	0	0
Labrum changes:		
<i>none</i>	17 (18%)	9 (10%)
<i>mild</i>	21 (23%)	18 (20%)
<i>moderate</i>	17 (18%)	19 (21%)
<i>severe</i>	32 (35%)	33 (36%)
<i>not accessible</i>	5* (5%)	13* (14%)
Iliopsoas bursitis:		
<i>absent</i>	89 (97%)	88 (96%)
<i>present</i>	3 (3%)	4 (4%)
<i>not accessible</i>	0	0
Trochanter bursitis:		
<i>absent</i>	90 (98%)	86 (94%)
<i>present</i>	2 (2%)	5 (6%)
<i>not accessible</i>	0	0
OA grading:		
<i>none</i>	32 (35%)	27 (29%)
<i>mild</i>	27 (29%)	35 (38%)
<i>moderate</i>	25 (27%)	22 (24%)
<i>severe</i>	8 (9%)	8 (9%)
<i>not accessible</i>	0	0
	Mean distance Operator A	Mean distance Operator B
Bone-capsule distance (incl. capsule) (mm):	6.7 (SD 2.5)	6.4 (SD 2.2)
Bone-capsule distance (excl. 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 visualize satisfactorily.

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&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 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.

***KLG; Kellgren-Lawrence Grade.

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 κ 0.4 - 0.7 (Table 3). Trochanter bursitis had $\kappa=0.3$ and iliopsoas bursitis $\kappa=0$.

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

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 (0.58-0.85)	
Bone-capsule distance (incl. capsule)	98%	87%		0,88 (0.82-0.93)
Bone-capsule distance (excl. capsule)	98%	88%		0.89 (0.84-0.93)
Overall joint effusion/synovitis	97 %	68%	0.90 (0.79-1.00)	
Femoral osteophytes	93%	82%	0.59 (0.45-0.70)	
Acetabular osteophytes	91%	85%	0.39 (0.20- 0.54)	
Femoral head deformation	95%	85%	0.70 (0.54- 0.82)	
Femoral cartilage changes	94%	83%	0.64 (0.51- 0.75)	
Femoral cartilage thickness*	96%	91%		0.59 (0.37-0.73)
Labrum changes**	89%	73%	0.60 (0.43-0.73)	
Iliopsoas bursitis	92%	93%	-0.04 (-0.09-0.01)	
Trochanter bursitis	95%	93%	0.26 (-0.18-0.70)	
OA grading	96%	80%	0.80 (0.71-0.87)	

* 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.

Table 4. Intra-rater reliability and agreement on ultrasound findings* assessed with a 4-6 weeks 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 – 0.86)	
Bone-capsule distance (incl. capsule)	100%	88%		0.99 (0.98 – 0.99)
Bone-capsule distance (excl. capsule)	100%	88%		0.99 (0.99 – 0.99)
Overall joint effusion/synovitis	96%	67%	0.87 (0.74 – 0.99)	
Femoral osteophytes	95%	75%	0.81 (0.72 – 0.88)	

Femoral cartilage thickness**	99%	93%		0.93 (0.88 – 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.

The mean difference between the operators on numeric measures was 0.3 mm (95% CI 0.1 – 0.6) for BCD (outer edge of the joint capsule) and -0,1 mm (95% CI -0.17 – -0.05) for cartilage thickness. Bland-Altman plots for these measures (Fig. 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 were $\kappa=0.5$ for both operators. For femoral and acetabular osteophytes, κ ranged from 0.2 to 0.4 (Table 5).

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

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

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 similar, with excellent inter- and intra-rater 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- and intra-rater 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- and intra-rater 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 has 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 radiologic 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 was our primary interest. In the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 intra-rater reliability is usually higher and image acquisition was eliminated as a source of variation, since intra-rater 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 (κ 0.2-0.5) as expected when comparing two different modalities. In relation to the grade of OA in general, we found moderate agreement (κ 0.5). While ultrasound can visualize inflammatory changes 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 Department's 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 Contributions 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 intra-rater 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.

Acknowledgements The authors would like to acknowledge Lise Bolander Malvang, science radiographer, Karin Kronborg Andersen, coordinating specialist in conventional radiography at the radiology department at 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 OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark.

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.

Data sharing statement No additional data are available.

Legends:

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

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Glyn-Jones S, Palmer AJR, Agricola R et al. Osteoarthritis. The Lancet 2015; 386: 376-387. doi:[https://doi.org/10.1016/S0140-6736\(14\)60802-3](https://doi.org/10.1016/S0140-6736(14)60802-3)

2. 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. doi:10.1055/s-0039-1681050

3. Hunter DJ, Arden N, Conaghan PG et al. Definition of osteoarthritis on MRI: results of a Delphi exercise. Osteoarthritis Cartilage 2011; 19: 963-969. doi:10.1016/j.joca.2011.04.017

4. Iagnocco A, Naredo E. Ultrasound of the osteoarthritic joint. Clin Exp Rheumatol 2017; 35: 527-534

5. Sudula SN. Imaging the hip joint in osteoarthritis: A place for ultrasound? Ultrasound 2016; 24: 111-118. doi:10.1177/1742271X16643118

6. Oo WM, Linklater JM, Daniel M et al. Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. Osteoarthritis Cartilage 2018. doi:10.1016/j.joca.2018.01.021. doi:10.1016/j.joca.2018.01.021

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: 1352-1356. doi:10.1016/j.ultrasmedbio.2012.04.006

8. Poulsen E, Overgaard S, Vestergaard JT et al. Pain distribution in primary care patients with hip osteoarthritis. Fam Pract 2016; 33: 601-606. doi:10.1093/fampra/cmw071

9. Kottner J, Audige L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64: 96-106. doi:10.1016/j.jclinepi.2010.03.002

10. Henrica C. W. de Vet CBT, Lidwine B. Mokkink and Dirk L. Knol. Measurement in medicine (Practical guides to Biostatistics And Epidemiology: Cambridge University Press; 2011

11. Klassbo 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. Moller I, Janta I, Backhaus M et al. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. Ann Rheum Dis 2017; 76: 1974-1979. doi:10.1136/annrheumdis-2017-211585

13. [Anonym]. The Minimum Training Recommendations for the Practice of Medical Ultrasound. Eur J Ultrasound; 29: 94-96

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. doi:10.3899/jrheum.181095. doi:10.3899/jrheum.181095

15. Carlo Martinoli SB. Ultrasound of the Musculoskeletal System: Springer-Verlag Berlin Heidelberg; 2007. doi:10.1007/978-3-540-28163-4

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: 1613-1619. doi:10.1136/ard.2005.050690

17. Wunderbaldinger P, Bremer C, Schellenberger E et al. Imaging features of iliopsoas bursitis. *Eur Radiol* 2002; 12: 409-415. doi:10.1007/s003300101041
18. Koski JM, Anttila PJ, Isomaki HA. Ultrasonography of the adult hip joint. *Scand J Rheumatol* 1989; 18: 113-117
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: 1141-1152. doi:10.1016/j.ultrasmedbio.2017.01.021
20. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage* 2007; 15 Suppl A: A1-56. doi:10.1016/j.joca.2006.11.009
21. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *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: 155-163. doi:10.1016/j.jcm.2016.02.012
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-174
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: 227-237. doi:10.1007/s00256-018-3010-9
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: 439-447
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: 314-318. doi:10.3109/17453670902988402
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. doi:10.1186/1471-2474-15-162
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: 1083-1086. doi:10.2214/ajr.12.10038

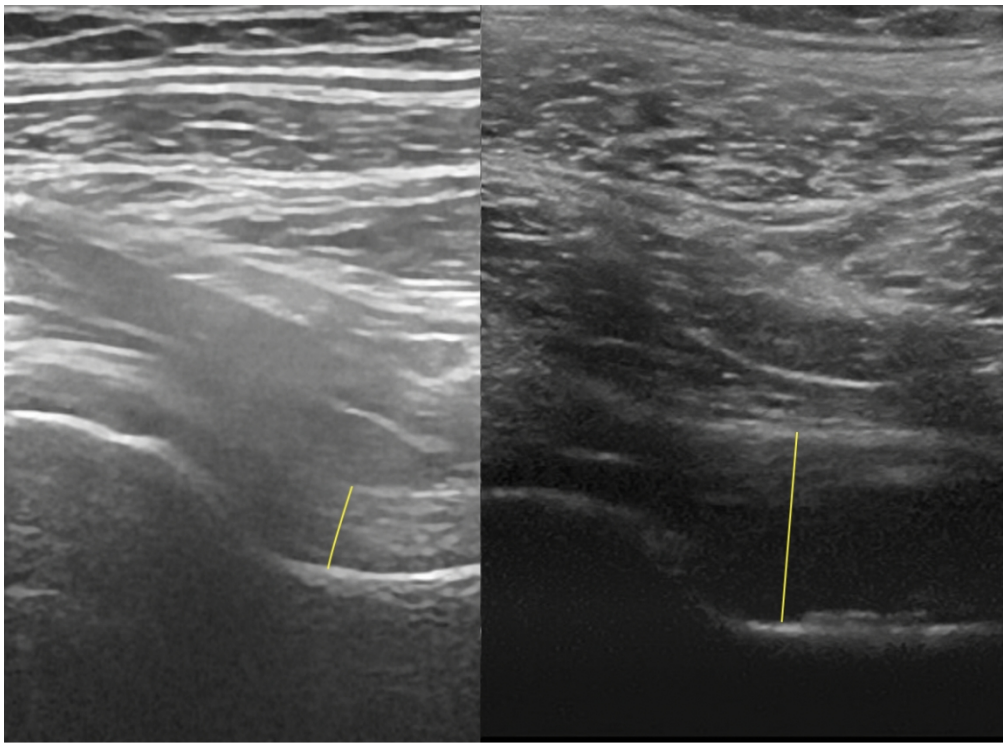


Figure 1 Anterior hip joint recess, longitudinal scan. The yellow line marks the bone-capsule distance. The hip on the right has effusion/synovitis.

100x74mm (300 x 300 DPI)

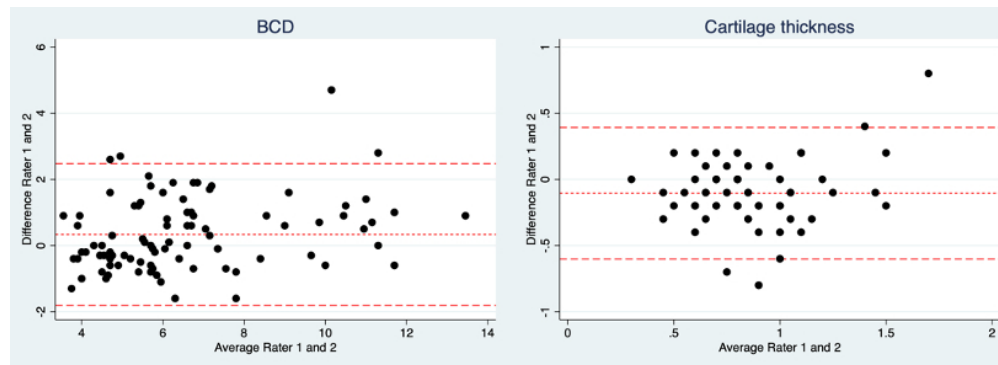


Figure 2. Bland-Altman plot with 95% limits of agreement for the two operator's recordings of bone-capsule distance (BCD) and cartilage thickness.

68x24mm (300 x 300 DPI)

Supplementary file 1. Ultrasound findings in the hip, scoring systems and definitions used in the assessment.

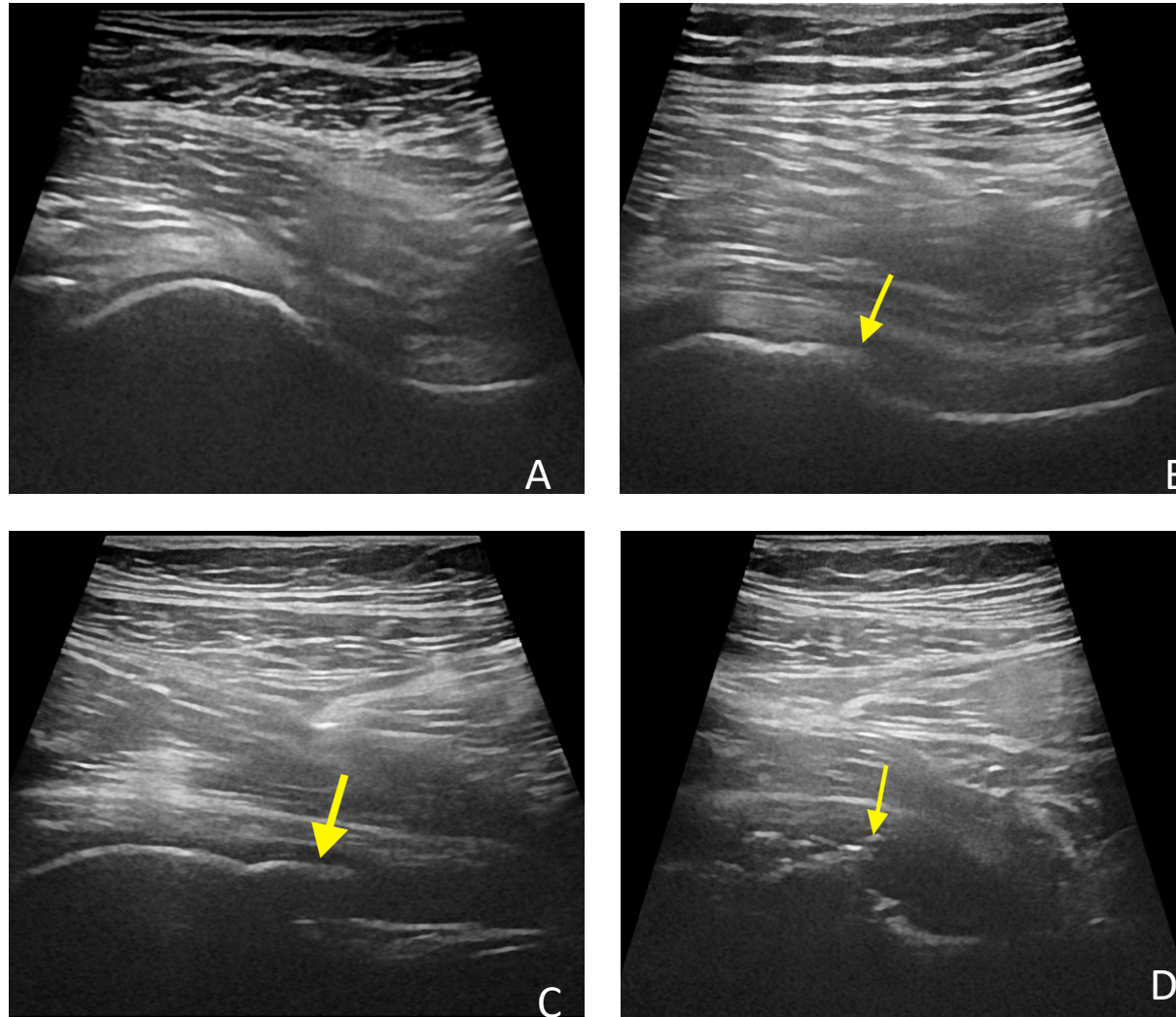
Ultrasound finding	Scoring system <i>Binary, ordinal (semi-quantitatively) or numerical</i>	Definition
Joint recess profile	<i>(ordinal)</i> Concave Straight Convex	The course of the anterior joint recess along the anterior surface of the femoral neck
Bone-capsule distance (BCD) in the anterior joint recess	<i>(numerical)</i> Number in mm.	Distance from the outer femoral cortex to the outer edge of the capsule. A second measure was made from the femoral cortex to the inside edge of the capsule
Overall joint effusion/synovitis	<i>(binary)</i> Present / absent	Overall assessment of all the ultrasound findings related to effusion / synovitis
Femoral osteophytes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Acetabular osteophytes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Femoral head deformation	<i>(ordinal)</i> None, mild, moderate or severe	Normal (round), mild (slightly flattened), moderate (very flattened), severe (no obvious contour or the femoral head can be defined)
Femoral cartilage changes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Femoral cartilage thickness	<i>(numerical)</i> Number in mm.	Anterior surface of femoral head, as close to labrum as possible.
Labrum changes	<i>(ordinal)</i> None, mild, moderate or severe	Normal (homogeneous echogenicity), mild (heterogeneous echogenicity and labrum poorly defined), moderate (definite pathology) and severe (pathology or degeneration to a degree where labrum was not defined)
Iliopsoas bursitis	<i>(binary)</i> Present / absent	Fluid associated with the iliopsoas tendon ventral to the hip joint
Trochanter bursitis	<i>(ordinal)</i> Present / absent	Fluid in any bursa in the trochanter region

OA grading	(ordinal) None, mild, moderate or severe	The degree of hip OA based on the ultrasound findings
------------	--	--

For peer review only

Supplementary file 2

Image examples



Supplementary Figure 1

Different degrees of osteoarthritis (OA) changes. All images are longitudinal scans of the anterior hip.

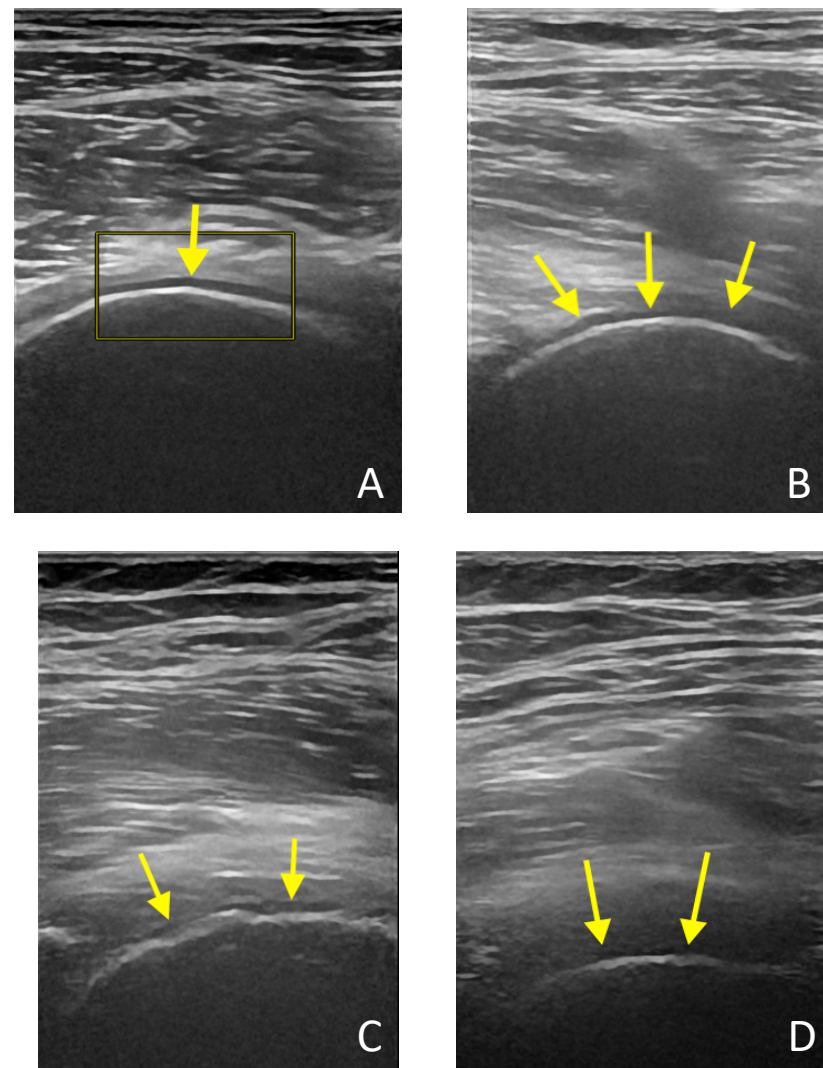
Figure 1A shows **No OA changes**.

Normal shape of femoral caput. Normal cartilage. No osteophytes. No effusion.

Figure 1B shows **Mild OA**: Small osteophyte (arrow). Normal shape of caput.

Figure 1C shows **Moderate OA**: Large osteophyte (arrow). Normal shape of caput.

Figure 1D shows **Severe OA**: Large osteophyte (arrow). Effusion in the hip joint. Severe deformation of caput femoris. The cartilage can't be defined.



Supplementary Figure 2

Cartilage changes. All images are longitudinal scans of the anterior hip.

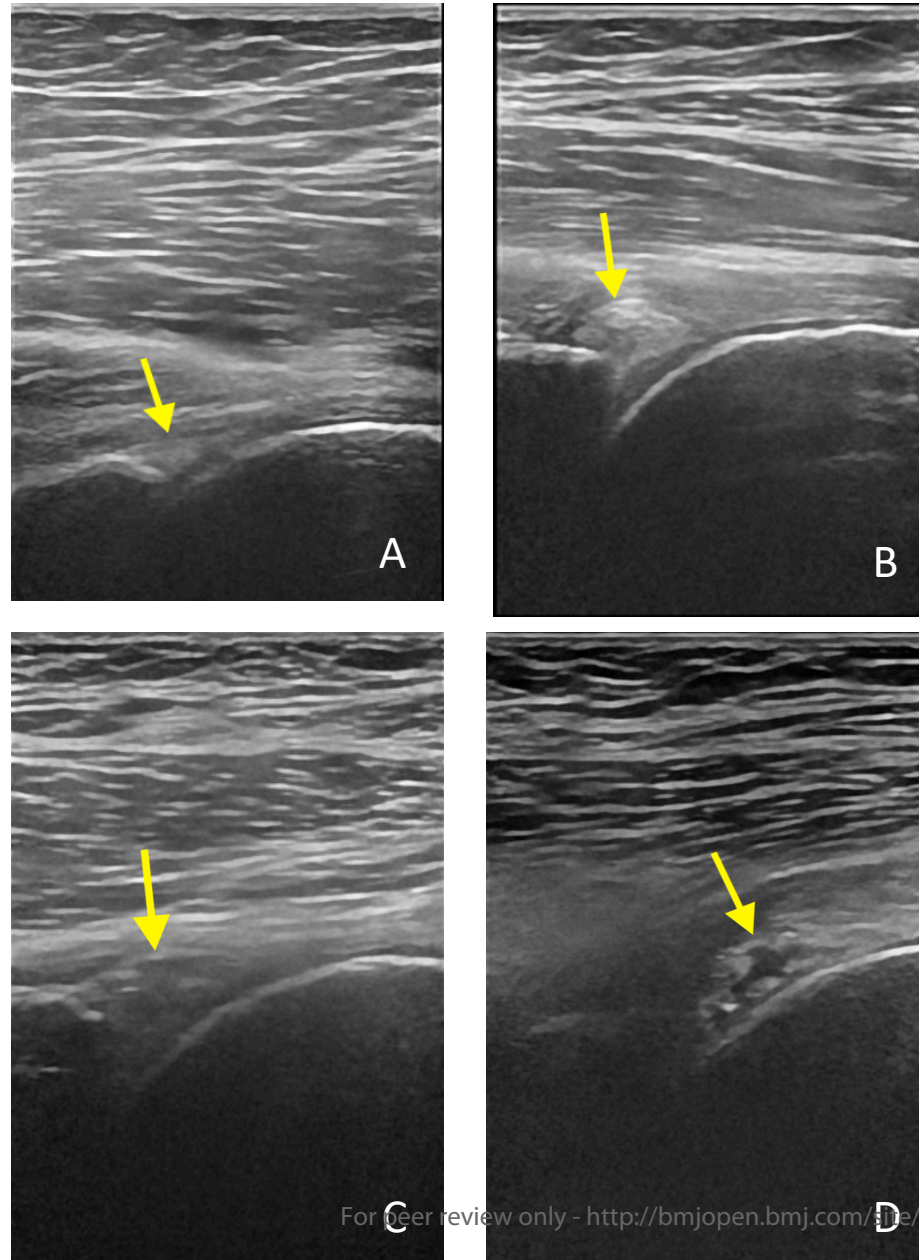
Figure 2A shows **Normal cartilage** (arrow).

Figure 2B shows **Mild cartilage change** (arrows).

Figure 2C shows **Moderate cartilage change** (arrows).

Figure 2D shows **Severe cartilage change** (arrows).

Cartilage can't be defined.



Supplementary Figure 3

Different degrees of **labral changes** (arrow). All images are longitudinal scans of the anterior hip.

Figure 3A shows **No changes**.

Figure 3B shows **Mild changes**.

Figure 3C shows **Moderate changes**.

Figure 3D shows **Severe changes**.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Kottner J, Audige L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64: 96-106.

Guidelines for Reporting Reliability and Agreement Studies (GRRAS).		Page
TITLE AND ABSTRACT	1. Identify in title or abstract that inter-rater/intra-rater reliability or agreement was investigated.	1
INTRODUCTION	2. Name and describe the diagnostic or measurement device of interest explicitly.	4
	3. Specify the subject population of interest.	4
	4. Specify the rater population of interest (if applicable).	NA
	5. Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	4
METHODS	6. Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	5
	7. Describe the sampling method.	5
	8. Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	5 6
	9. State whether measurements/ratings were conducted independently.	5
	10. Describe the statistical analysis.	7
RESULTS	11. State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	5 7
	12. Describe the sample characteristics of raters and subjects (e.g. training, experience).	5 7-9
	13. Report estimates of reliability and agreement including measures of statistical uncertainty.	10
DISCUSSION	14. Discuss the practical relevance of results.	12
AUXILIARY MATERIAL	15. Provide detailed results if possible (e.g. online)	NA

BMJ Open

Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis. An inter- and intra-rater reliability study.

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2020-038643.R2
Article Type:	Original research
Date Submitted by the Author:	05-Oct-2020
Complete List of Authors:	Clausen, Stine; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Lillebaelt Hospital Vejle, Department of Radiology Kjær, Søren; Silkeborg Regional Hospital, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways Fredberg, Ulrich; Silkeborg Regional Hospital, Diagnostic Centre, University Research Clinic for Innovative Patient Pathways; Odense University Hospital, The Rheumatology Research Unit Terslev, Lene; Rigshospitalet Glostrup, Copenhagen Center for Arthritis Research and Center for Rheumatology and Spine Diseases; University of Copenhagen Faculty of Health and Medical Sciences, Department of Clinical Medicine Hartvigsen, Jan; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Nordic Institute of Chiropractic and Clinical Biomechanics Arnbak, Bodil; University of Southern Denmark Faculty of Health Sciences, Department of Sports Science and Clinical Biomechanics; Lillebaelt Hospital Vejle, Department of Radiology
Primary Subject Heading:	Radiology and imaging
Secondary Subject Heading:	Rheumatology
Keywords:	Ultrasound < RADIOLOGY & IMAGING, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Diagnostic radiology < RADIOLOGY & IMAGING, RHEUMATOLOGY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Ultrasound imaging in patients with hip pain and suspected hip osteoarthritis. An inter- and intra-rater reliability study.

Stine Clausen^{1,2}
Søren Geil Kjær³
Ulrich Fredberg^{3,4}
Lene Terslev^{5,6}
Jan Hartvigsen^{1,7}
Bodil Arnbak^{1,2}

Corresponding author: Stine Clausen, University of Southern Denmark, Campusvej 55, 5230 Odense M, sclausen@health.sdu.dk telephone +45 30256613 Fax: none

¹ Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark.

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

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

⁴The Rheumatology Research Unit, Odense University Hospital, University of Southern Denmark, Denmark.

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

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

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

Keywords: Ultrasound, Hip, Osteoarthritis, Reliability

Word count, excluding title page, abstract, references, figures and tables: 2948

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

Objectives: The objectives of this study were to assess 1) inter- and intra-rater reliability of ultrasound imaging in patients with hip osteoarthritis, and 2) agreement between ultrasound and x-ray findings of hip osteoarthritis using validated OMERACT ultrasound definitions for pathology.

Design: An inter-rater and intra-rater 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- and intra-rater reliability and agreement between ultrasound imaging and x-ray were assessed using Cohen's ordinal kappa statistics for binominal categorical variables and weighted kappa for ordered categorical variables.

Results: Kappa values (κ) for inter-rater reliability was 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 0.4-0.7. Intra-rater reliability had κ equal or higher compared to inter-rater reliability. Agreement between ultrasound and x-ray findings ranged from κ 0.2-0.5.

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.

Strengths and limitations of this study

- Using OMERACT 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 OA, making the results transferable to other clinical settings.
- Intra-rater 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's guidelines, and variation in acquisition procedures might have affected the findings.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

INTRODUCTION

Osteoarthritis (OA) is characterized by progressive destruction of articular cartilage, changes in bone tissue, osteophyte formation and joint inflammation resulting in loss of normal joint function[1]. The pathophysiological changes can be visualized 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 to visualize 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 modalities[3]; 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 epidemiologic 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 is needed.

The primary objective of this study was to assess the inter- 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 radiologic findings related to hip OA.

MATERIALS AND METHODS

Guidelines for Reporting Reliability and Agreement Studies[9] 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 a 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 minutes for each examiner.

The ultrasound operators were a chiropractor and a rheumatologist. Both had 10-15 year 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 Reumatology (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 were assessed and rated semi-quantitatively (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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 Supplementary file 1. Image examples are illustrated in Supplementary file 2.

Ultrasound examination 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 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 (Fig. 1). We measured BCD from the cortical surface of the femoral neck to both the inner and the outer edge of the joint capsule, the latter combining joint fluid and synovium/capsule. BCD increase to 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 hypoechoic 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 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 grading was the operator's overall assessment of the findings mentioned above. Image examples of the different stages are illustrated in Supplementary file 2.

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

Where a rating was questionable, the finding was rated in the lowest category in question. Representative images were stored during the examinations. After 4 to 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, bone-capsule distance, 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's guidelines) images were recorded standing, unless the patient could not stand correctly (13 hips were recorded lying). An imaging specialist with ten years' 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 radiologic hip OA was assessed using Kellgren-Lawrence grading system (KLG)[21]. Radiographic hip osteoarthritis was defined as a KLG ≥ 2 .

Statistical analysis

The statistical analysis was performed using STATA/IC 15.1. Inter- and intra-rater reliability and agreement between ultrasound imaging and x-ray were assessed using Cohen's ordinal kappa statistics for binominal categorical variables and weighted kappa for ordered categorical variables. Quadratic weights were applied according to the number of categories and a 95% confidence interval (CI) was calculated by bootstrap resampling with 1000 repetitions for ordered categorical variables. The interclass correlation coefficient for agreement (ICC, 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 $\pm 1.96 \times$ 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 ($0.0 \leq \kappa \leq 0.2$), fair ($0.2 < \kappa \leq 0.4$), moderate ($0.4 < \kappa \leq 0.6$), substantial ($0.6 < \kappa \leq 0.8$) and almost perfect ($0.8 < \kappa \leq 1$)[24]. The ICC for agreement was interpreted as follows: ICC < 0.5 = poor, $0.5 \leq \text{ICC} \leq 0.75$ = moderate, $0.75 < \text{ICC} \leq 0.9$ = good and > 0.9 = excellent[22].

Ethics

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).

Patient and Public Involvement

1
2
3
4 Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans
5 of our research.
6
7

8
9 **RESULTS**

10
11 Bilateral hips in 50 participants (n=92) were included in the study. Due to previous surgery, eight
12 non-painful hips were excluded. Of the included participants, 43 were referred from general
13 practitioners and 7 from orthopaedic surgeons, 32 (64%) were women and 26 (52%) had symptoms
14 for more than 16 weeks. Age ranged from 42 to 90 years (median 67 years), mean BMI was 26.9
15 (range 18.4 - 36.6). Mean HOOS on pain and function in daily living were 49 (SD 19) and 53 (SD
16 19), respectively (100=normal function). Because we followed Department guidelines all participants
17 had an x-ray of the painful hip, but only some had bilateral hip (pelvic AP) resulting in 63 hip x-rays.
18 Of these, 36 (57%) had KLG 2 or more. This is defined as radiological OA. On an individual level
19 28 of the 50 participants had radiographic OA in either one or both hips.
20
21 Prevalence of ultrasound and x-ray findings are shown in Tables 1 and 2, respectively. The most
22 prevalent ultrasound finding was labrum changes (53-57% had moderate or severe changes) and least
23 prevalent findings were effusion in iliopsoas or trochanter bursas (2-6%).
24
25
26
27
28
29
30
31
32
33

34 Table 1. Prevalence and mean measure of ultrasound findings in 50 participants (92 hips) examined
35 by two operators blinded to each other's findings.
36

Ultrasound finding	Prevalence Operator A	Prevalence Operator B
Joint recess profile:		
<i>concave</i>	71 (77%)	73 (79%)
<i>straight</i>	12 (13%)	10 (11%)
<i>convex</i>	9 (10%)	9 (10%)
<i>not accessible</i>	0	0
Overall joint effusion/synovitis:		
<i>absent</i>	74 (80%)	73 (79%)
<i>present</i>	18 (20%)	19 (21%)
<i>not accessible</i>	0	0
Femoral osteophytes:		
<i>none</i>	23 (25%)	18 (20%)
<i>mild</i>	41 (45%)	43 (47%)
<i>moderate</i>	19 (21%)	22 (24%)
<i>severe</i>	9 (10%)	9 (10%)
<i>not accessible</i>	0	0
Acetabular osteophytes:		
<i>none</i>	38 (41%)	40 (44%)

<i>mild</i>	31 (34%)	38 (42%)
<i>moderate</i>	18 (20%)	11 (12%)
<i>severe</i>	5 (5%)	1 (1%)
<i>not accessible</i>	0	2 (1%)
Femoral head deformation:		
<i>none</i>	57 (62%)	49 (53%)
<i>mild</i>	23 (25%)	29 (32%)
<i>moderate</i>	8 (9%)	10 (11%)
<i>severe</i>	4 (4%)	4 (4%)
<i>not accessible</i>	0	0
Femoral cartilage changes:		
<i>none</i>	33 (36%)	25 (27%)
<i>mild</i>	31 (34%)	36 (39%)
<i>moderate</i>	23 (25%)	27 (29%)
<i>severe</i>	5 (5%)	4 (4%)
<i>not accessible</i>	0	0
Labrum changes:		
<i>none</i>	17 (18%)	9 (10%)
<i>mild</i>	21 (23%)	18 (20%)
<i>moderate</i>	17 (18%)	19 (21%)
<i>severe</i>	32 (35%)	33 (36%)
<i>not accessible</i>	5* (5%)	13* (14%)
Iliopsoas bursitis:		
<i>absent</i>	89 (97%)	88 (96%)
<i>present</i>	3 (3%)	4 (4%)
<i>not accessible</i>	0	0
Trochanter bursitis:		
<i>absent</i>	90 (98%)	86 (94%)
<i>present</i>	2 (2%)	5 (6%)
<i>not accessible</i>	0	0
OA grading:		
<i>none</i>	32 (35%)	27 (29%)
<i>mild</i>	27 (29%)	35 (38%)
<i>moderate</i>	25 (27%)	22 (24%)
<i>severe</i>	8 (9%)	8 (9%)
<i>not accessible</i>	0	0
	Mean distance Operator A	Mean distance Operator B
Bone-capsule distance (incl. capsule) (mm):	6.7 (SD 2.5)	6.4 (SD 2.2)
Bone-capsule distance (excl. 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 visualize satisfactorily.

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&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 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.

***KLG; Kellgren-Lawrence Grade.

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 κ 0.4 - 0.7 (Table 3). Trochanter bursitis had $\kappa=0.3$ and iliopsoas bursitis $\kappa=0$.

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

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 (0.58-0.85)	
Bone-capsule distance (incl. capsule)	98%	87%		0.88 (0.82-0.93)
Bone-capsule distance (excl. capsule)	98%	88%		0.89 (0.84-0.93)
Overall joint effusion/synovitis	97 %	68%	0.90 (0.79-1.00)	
Femoral osteophytes	93%	82%	0.59 (0.45-0.70)	
Acetabular osteophytes	91%	85%	0.39 (0.20- 0.54)	
Femoral head deformation	95%	85%	0.70 (0.54- 0.82)	
Femoral cartilage changes	94%	83%	0.64 (0.51- 0.75)	
Femoral cartilage thickness*	96%	91%		0.59 (0.37-0.73)
Labrum changes**	89%	73%	0.60 (0.43-0.73)	
Iliopsoas bursitis	92%	93%	-0.04 (-0.09-0.01)	
Trochanter bursitis	95%	93%	0.26 (-0.18-0.70)	
OA grading	96%	80%	0.80 (0.71-0.87)	

* 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.

Table 4. Intra-rater reliability and agreement on ultrasound findings* assessed with a 4-6 weeks 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 – 0.86)	
Bone-capsule distance (incl. capsule)	100%	88%		0.99 (0.98 – 0.99)
Bone-capsule distance (excl. capsule)	100%	88%		0.99 (0.99 – 0.99)
Overall joint effusion/synovitis	96%	67%	0.87 (0.74 – 0.99)	
Femoral osteophytes	95%	75%	0.81 (0.72 – 0.88)	

Femoral cartilage thickness**	99%	93%		0.93 (0.88 – 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.

The mean difference between the operators on numeric measures was 0.3 mm (95% CI 0.1 – 0.6) for BCD (outer edge of the joint capsule) and -0,1 mm (95% CI -0.17 – -0.05) for cartilage thickness. Bland-Altman plots for these measures (Fig. 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 were $\kappa=0.5$ for both operators. For femoral and acetabular osteophytes, κ ranged from 0.2 to 0.4 (Table 5).

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

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

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 similar, with excellent inter- and intra-rater 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- and intra-rater 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- and intra-rater 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 has 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 radiologic 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 was our primary interest. In the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 intra-rater reliability is usually higher and image acquisition was eliminated as a source of variation, since intra-rater 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 (κ 0.2-0.5) as expected when comparing two different modalities. In relation to the grade of OA in general, we found moderate agreement (κ 0.5). While ultrasound can visualize inflammatory changes 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 Department's 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 Contributions 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 intra-rater 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.

Acknowledgements The authors would like to acknowledge Lise Bolander Malvang, science radiographer, Karin Kronborg Andersen, coordinating specialist in conventional radiography at the radiology department at 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 OPEN, Odense Patient data Explorative Network, Odense University Hospital, Odense, Denmark.

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.

Data sharing statement No additional data are available.

Legends:

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

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

1. Glyn-Jones S, Palmer AJR, Agricola R et al. Osteoarthritis. *The Lancet* 2015; 386: 376-387. doi:[https://doi.org/10.1016/S0140-6736\(14\)60802-3](https://doi.org/10.1016/S0140-6736(14)60802-3)

2. 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. doi:10.1055/s-0039-1681050

3. Hunter DJ, Arden N, Conaghan PG et al. Definition of osteoarthritis on MRI: results of a Delphi exercise. *Osteoarthritis Cartilage* 2011; 19: 963-969. doi:10.1016/j.joca.2011.04.017

4. Iagnocco A, Naredo E. Ultrasound of the osteoarthritic joint. *Clin Exp Rheumatol* 2017; 35: 527-534

5. Sudula SN. Imaging the hip joint in osteoarthritis: A place for ultrasound? *Ultrasound* 2016; 24: 111-118. doi:10.1177/1742271X16643118

6. Oo WM, Linklater JM, Daniel M et al. Clinimetrics of ultrasound pathologies in osteoarthritis: systematic literature review and meta-analysis. *Osteoarthritis Cartilage* 2018. doi:10.1016/j.joca.2018.01.021. doi:10.1016/j.joca.2018.01.021

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: 1352-1356. doi:10.1016/j.ultrasmedbio.2012.04.006

8. Poulsen E, Overgaard S, Vestergaard JT et al. Pain distribution in primary care patients with hip osteoarthritis. *Fam Pract* 2016; 33: 601-606. doi:10.1093/fampra/cmw071

9. Kottner J, Audige L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011; 64: 96-106. doi:10.1016/j.jclinepi.2010.03.002

10. Henrica C. W. de Vet CBT, Lidwine B. Mokkink and Dirk L. Knol. Measurement in medicine (Practical guides to Biostatistics And Epidemiology: Cambridge University Press; 2011

11. Klassbo 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. Moller I, Janta I, Backhaus M et al. The 2017 EULAR standardised procedures for ultrasound imaging in rheumatology. *Ann Rheum Dis* 2017; 76: 1974-1979. doi:10.1136/annrheumdis-2017-211585

13. [Anonym]. The Minimum Training Recommendations for the Practice of Medical Ultrasound. *Eur J Ultrasound*; 29: 94-96

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. doi:10.3899/jrheum.181095. doi:10.3899/jrheum.181095

15. Carlo Martinoli SB. *Ultrasound of the Musculoskeletal System*: Springer-Verlag Berlin Heidelberg; 2007. doi:10.1007/978-3-540-28163-4

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: 1613-1619. doi:10.1136/ard.2005.050690

17. Wunderbaldinger P, Bremer C, Schellenberger E et al. Imaging features of iliopsoas bursitis. *Eur Radiol* 2002; 12: 409-415. doi:10.1007/s003300101041
18. Koski JM, Anttila PJ, Isomaki HA. Ultrasonography of the adult hip joint. *Scand J Rheumatol* 1989; 18: 113-117
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: 1141-1152. doi:10.1016/j.ultrasmedbio.2017.01.021
20. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. *Osteoarthritis Cartilage* 2007; 15 Suppl A: A1-56. doi:10.1016/j.joca.2006.11.009
21. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *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: 155-163. doi:10.1016/j.jcm.2016.02.012
23. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307-310
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159-174
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: 227-237. doi:10.1007/s00256-018-3010-9
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: 439-447
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: 314-318. doi:10.3109/17453670902988402
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. doi:10.1186/1471-2474-15-162
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: 1083-1086. doi:10.2214/ajr.12.10038

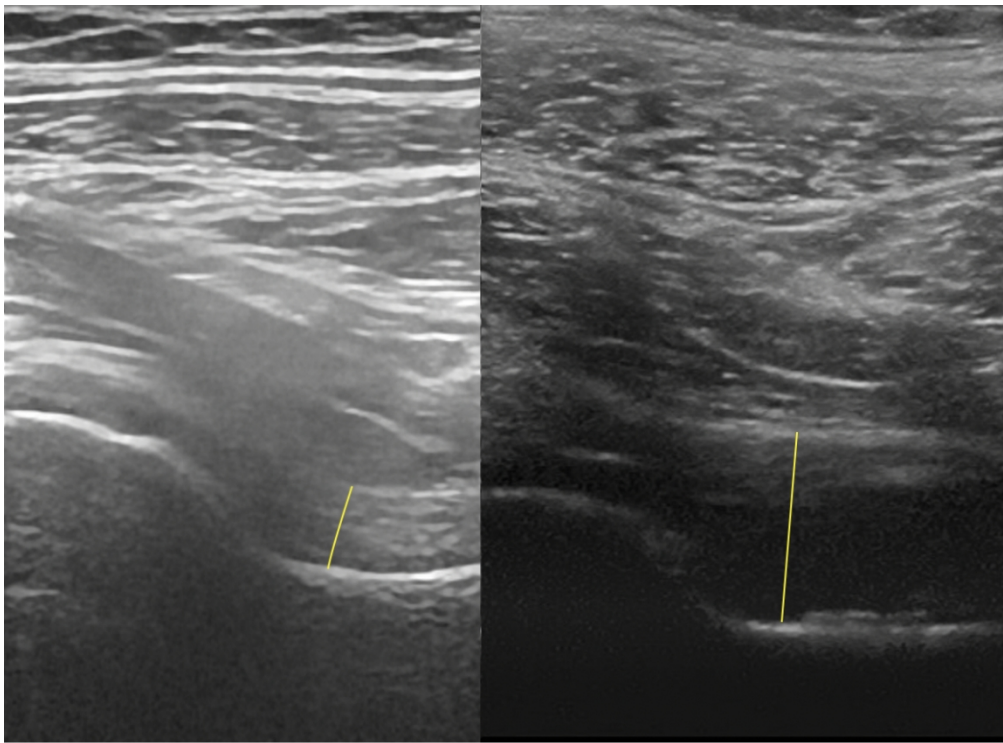


Figure 1 Anterior hip joint recess, longitudinal scan. The yellow line marks the bone-capsule distance. The hip on the right has effusion/synovitis.

100x74mm (300 x 300 DPI)

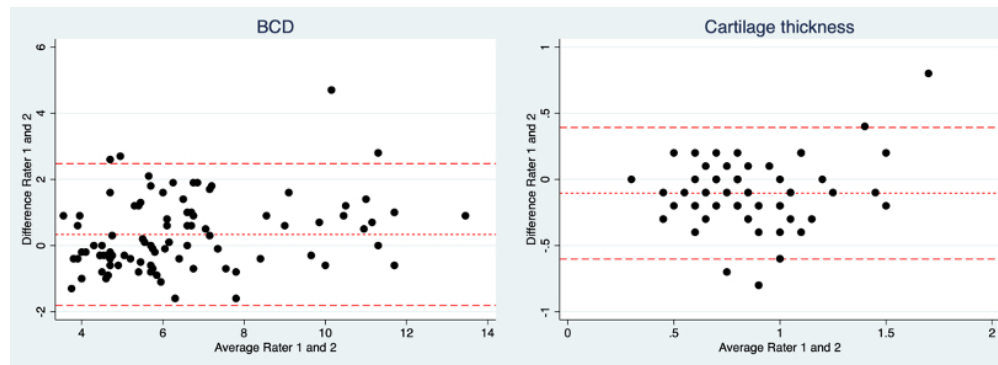


Figure 2. Bland-Altman plot with 95% limits of agreement for the two operator's recordings of bone-capsule distance (BCD) and cartilage thickness.

68x24mm (300 x 300 DPI)

Supplementary file 1. Ultrasound findings in the hip, scoring systems and definitions used in the assessment.

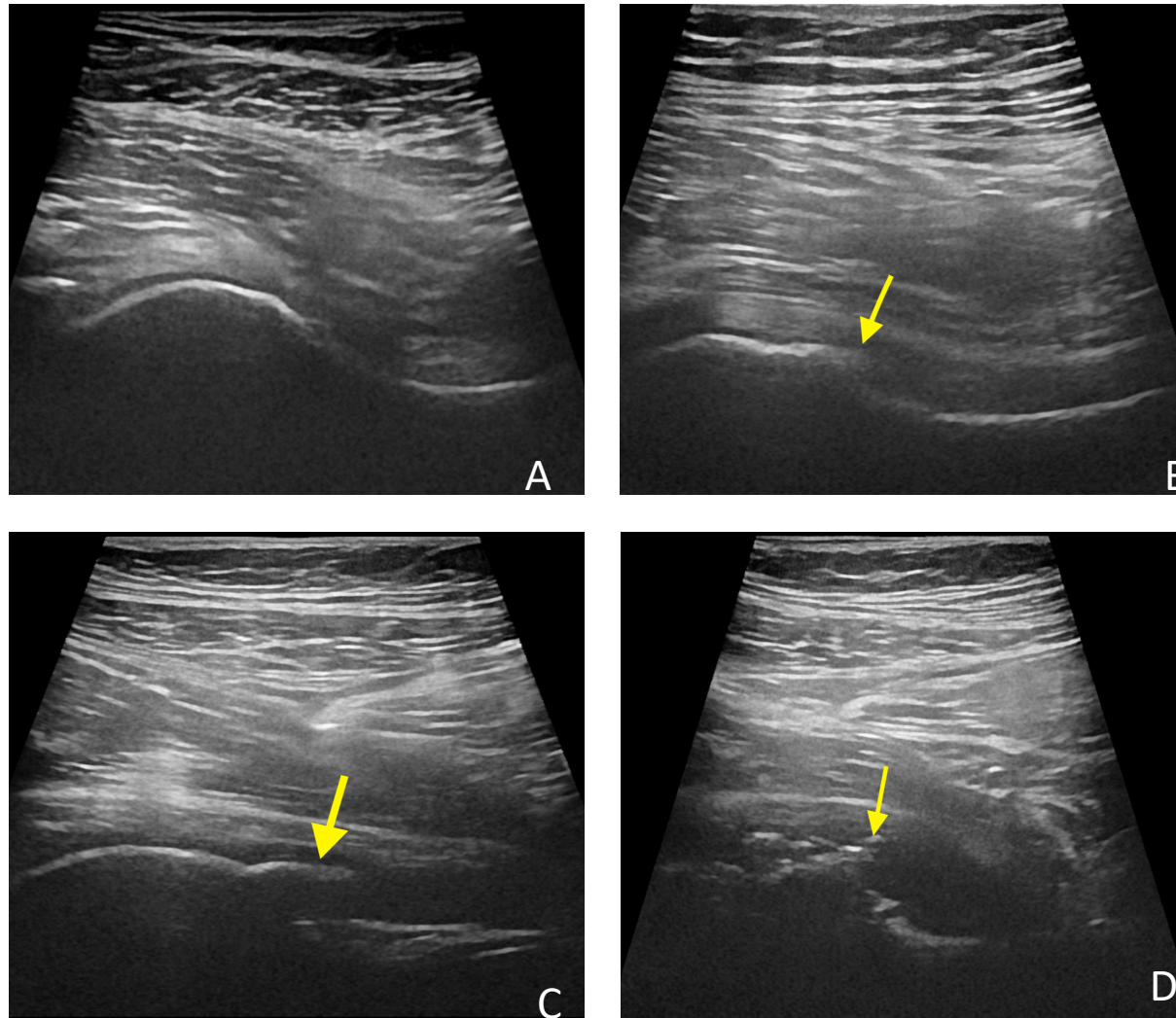
Ultrasound finding	Scoring system <i>Binary, ordinal (semi-quantitatively) or numerical</i>	Definition
Joint recess profile	<i>(ordinal)</i> Concave Straight Convex	The course of the anterior joint recess along the anterior surface of the femoral neck
Bone-capsule distance (BCD) in the anterior joint recess	<i>(numerical)</i> Number in mm.	Distance from the outer femoral cortex to the outer edge of the capsule. A second measure was made from the femoral cortex to the inside edge of the capsule
Overall joint effusion/synovitis	<i>(binary)</i> Present / absent	Overall assessment of all the ultrasound findings related to effusion / synovitis
Femoral osteophytes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Acetabular osteophytes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Femoral head deformation	<i>(ordinal)</i> None, mild, moderate or severe	Normal (round), mild (slightly flattened), moderate (very flattened), severe (no obvious contour or the femoral head can be defined)
Femoral cartilage changes	<i>(ordinal)</i> None, mild, moderate or severe	Classified according to OMERACT guidelines
Femoral cartilage thickness	<i>(numerical)</i> Number in mm.	Anterior surface of femoral head, as close to labrum as possible.
Labrum changes	<i>(ordinal)</i> None, mild, moderate or severe	Normal (homogeneous echogenicity), mild (heterogeneous echogenicity and labrum poorly defined), moderate (definite pathology) and severe (pathology or degeneration to a degree where labrum was not defined)
Iliopsoas bursitis	<i>(binary)</i> Present / absent	Fluid associated with the iliopsoas tendon ventral to the hip joint
Trochanter bursitis	<i>(ordinal)</i> Present / absent	Fluid in any bursa in the trochanter region

OA grading	(ordinal) None, mild, moderate or severe	The degree of hip OA based on the ultrasound findings
------------	--	--

For peer review only

Supplementary file 2

Image examples



Supplementary Figure 1

Different degrees of osteoarthritis (OA) changes. All images are longitudinal scans of the anterior hip.

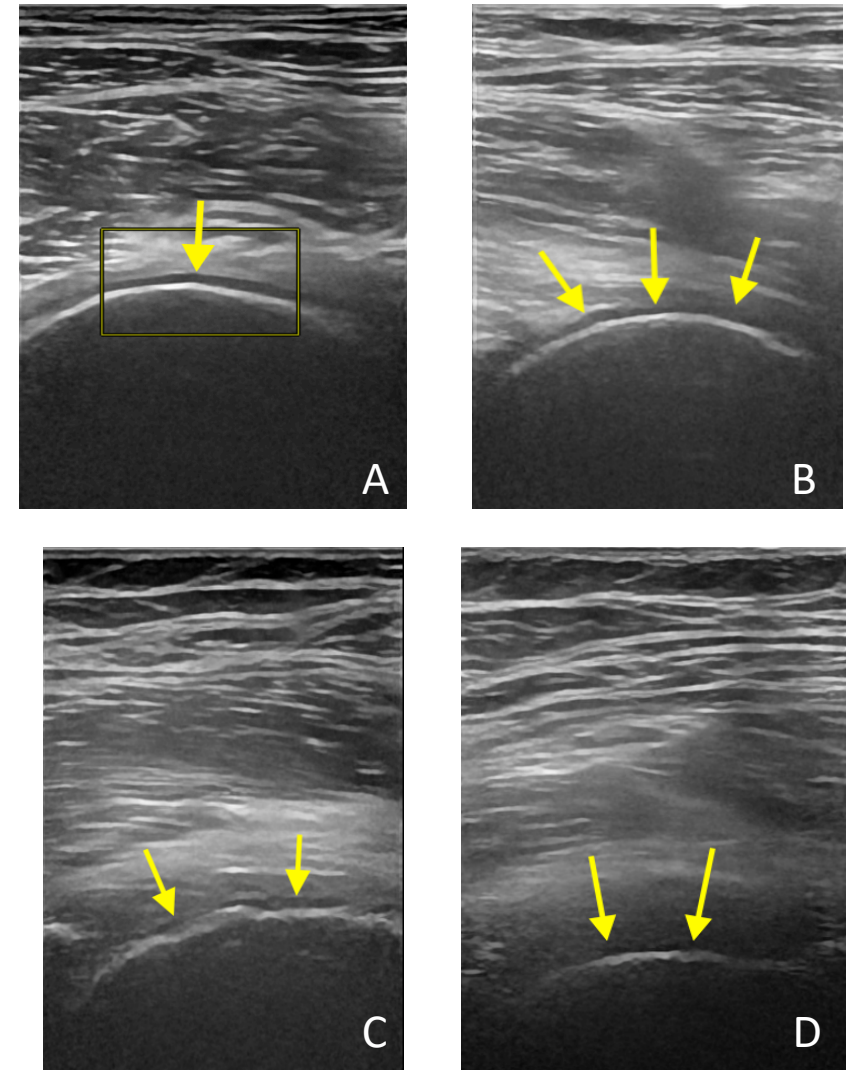
Figure 1A shows **No OA changes**.

Normal shape of femoral caput. Normal cartilage. No osteophytes. No effusion.

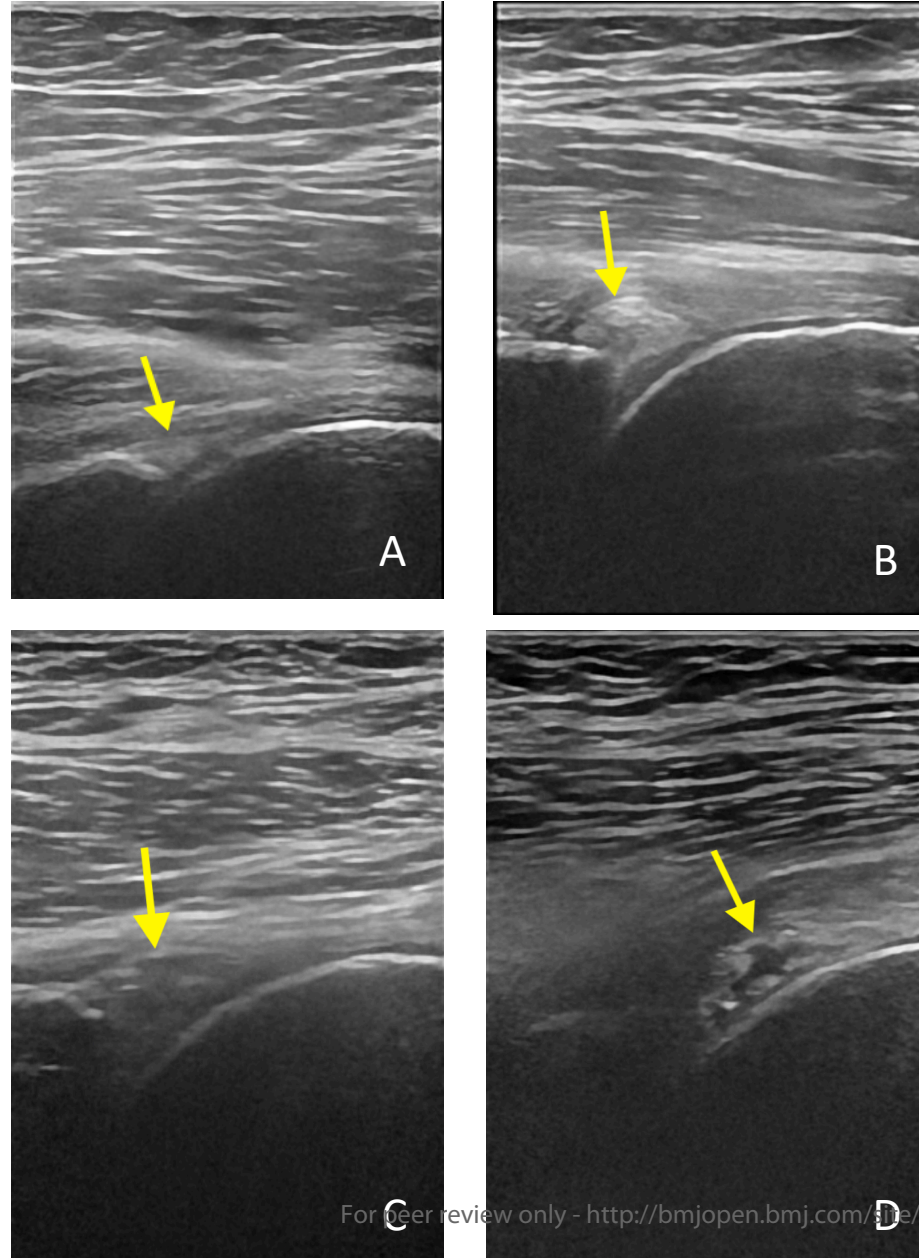
Figure 1B shows **Mild OA**: Small osteophyte (arrow). Normal shape of caput.

Figure 1C shows **Moderate OA**: Large osteophyte (arrow). Normal shape of caput.

Figure 1D shows **Severe OA** Large osteophyte (arrow). Effusion in the hip joint. Severe deformation of caput femoris. The cartilage can't be defined.



Supplementary Figure 2
Cartilage changes. All images are longitudinal scans of the anterior hip.
Figure 2A shows **Normal cartilage** (arrow).
Figure 2B shows **Mild cartilage change** (arrows).
Figure 2C shows **Moderate cartilage change** (arrows).
Figure 2D shows **Severe cartilage change** (arrows).
Cartilage can't be defined.



Supplementary Figure 3

Different degrees of **labral changes** (arrow). All images are longitudinal scans of the anterior hip.

Figure 3A shows **No changes**.

Figure 3B shows **Mild changes**.

Figure 3C shows **Moderate changes**.

Figure 3D shows **Severe changes**.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Kottner J, Audige L, Brorson S et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol 2011; 64: 96-106.

Guidelines for Reporting Reliability and Agreement Studies (GRRAS).		Page
TITLE AND ABSTRACT	1. Identify in title or abstract that inter-rater/intra-rater reliability or agreement was investigated.	1
INTRODUCTION	2. Name and describe the diagnostic or measurement device of interest explicitly.	4
	3. Specify the subject population of interest.	4
	4. Specify the rater population of interest (if applicable).	NA
	5. Describe what is already known about reliability and agreement and provide a rationale for the study (if applicable).	4
	6. Explain how the sample size was chosen. State the determined number of raters, subjects/objects, and replicate observations.	5
METHODS	7. Describe the sampling method.	5
	8. Describe the measurement/rating process (e.g. time interval between repeated measurements, availability of clinical information, blinding).	5 6
	9. State whether measurements/ratings were conducted independently.	5
	10. Describe the statistical analysis.	7
	11. State the actual number of raters and subjects/objects which were included and the number of replicate observations which were conducted.	5 7
RESULTS	12. Describe the sample characteristics of raters and subjects (e.g. training, experience).	5 7-9
	13. Report estimates of reliability and agreement including measures of statistical uncertainty.	10
	14. Discuss the practical relevance of results.	12
AUXILIARY MATERIAL	15. Provide detailed results if possible (e.g. online)	NA