The development of a preliminary ultrasonographic scoring system for features of hand osteoarthritis

H I Keen,[1] F Lavie,[2] R J Wakefield,[1] M-A D'Agostino,[3] H Berner Hammer,[4]
E M A Hensor,[1] A Pendleton,[5] D Kane,[6] H Guerini,[2] C Schueller-Weidekamm,[7]
M C Kortekaas,[8] F Birrel,[9] M Kloppenburg,[8] T Stamm,[10] I Watt,[8] J S Smolen,[7]
E Maheu,[11] M Dougados,[2] P G Conaghan[1]

[1] Academic Unit of Musculoskeletal Disease, University of Leeds, Leeds, UK; [2] Rheumatology B, Cochin Hospital, Paris France; [3] Ambroise Pare Hospital, Boulogne-Billancourt, France; [4] Diakonhjemmet Hospital, Oslo, Norway; [5] Musgrave Park Hospital Campus Belfast Health and Social Care Trust, UK; [6] Adelaide and Meath Hospital, Dublin, Ireland; [7] Medical University Hospital, Vienna, Austria; [8] Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands; [9] Newcastle University, Newcastle, UK; [10] Department of Internal Medicine III, Division of Rheumatology, Vienna Medical University, Vienna, Austria; [11] Saint-Antoine Hospital, Paris, France

Correspondence to:
Helen Keen, Academic Unit of Musculoskeletal Disease, Chapel Allerton Hospital, Chapeltown Road, Leeds LS7 4SA, UK; h.i. keen@leeds.ac.uk

Accepted 12 August 2007
Published Online First 17 August 2007

ABSTRACT

Objectives: Painful osteoarthritis (OA) of the hand is common and a validated ultrasound (US) scoring system would be valuable for epidemiological and therapeutic outcome studies. US is increasingly used to assess peripheral joints, though most of the US focus in rheumatic diseases has been on rheumatoid arthritis. We aimed to develop a preliminary US hand OA scoring system, initially focusing on relevant pathological features with potentially high reliability.

Methods: A group of experts in the fields of OA, US and novel tool development agreed on domains and suggested scaling of the items to be used in US hand OA scoring systems. A multi-observer reliability exercise was then performed to evaluate the draft items.

Results: Synovitis (grey scale and Power Doppler) and osteophytes (representing activity and damage domains) were included and evaluated as the initial components of the scoring system. All three features were evaluated for their presence/absence and if present were scored using a 1–3 scale. The reliability exercise demonstrated intra-reader κ values of 0.444–1.0, 0.211–1.0 and 0.087–1.0 for grey scale synovitis, power Doppler and osteophytes respectively. Inter-reader reliability κ values were 0.398, 0.327 and 0.530 grey-scale synovitis, power Doppler and osteophytes respectively. Without extensive standardisation, both intra- and inter-reader reliability were moderately good.

Conclusions: The draft scoring system demonstrated substantive to almost perfect percentage exact agreement on the presence/absence of the selected OA features and moderate to substantive percentage exact agreement on semi-quantitative grading. This preliminary process provides a good basis from which to further develop an US outcome tool for hand OA that has the potential to be utilised in multicentre clinical trials.

Osteoarthritis (OA) is the most common joint disease[1] and is associated with significant health economic consequences[2 3] The prevalence of radiographic OA has been well documented in epidemiological studies; however, the prevalence of symptomatic hand OA is not well documented.[4] The Framingham study has estimated the prevalence to be as high as 26% of women and 12% of men over 70.[4] While treatment recommendations focus on a holistic approach, the pharmaceutical options are currently largely limited to analgesics[5–7] However, the spectrum of pharmaceutical therapies is expanding, with a recent increase in interest in potential disease modifying therapies in OA.[8 9]

OA is assessed clinically with attention to symptoms and signs, and confirmed by structural changes on radiographs. Similarly, trials focus on clinical and structural outcomes. The Osteoarthritis Research Society International (OARSI) group has recently published guidelines for conduct of clinical trials of OA of the hand,[10] recommending conventional radiographs (CRs) as the standard for assessing structural outcomes. However, they acknowledged that other novel imaging techniques may play a part and require further validation.

Ultrasound (US) appears favourably placed to assess OA both in the clinic and in clinical trials. It has a higher resolution than CR, does not involve ionising radiation, and allows multi-planar, dynamic imaging of joints. In addition, recent studies in inflammatory arthritis have demonstrated US to be more sensitive to synovitis than clinical examination[11–13] more sensitive than CR to the presence of cortical defects,[14 15] and have reasonable sensitivity compared with magnetic resonance imaging for the presence of synovitis and cortical defects[13 16]

A group of experts in OA, US and outcome measures, met under the auspices of the Disease Characteristics in Hand OA Group (DICHOA), to gain consensus on the content, and take preliminary steps towards validation of an US scoring system for OA of the hand.

METHODS

Experts in OA, outcomes measures and ultrasonography from six countries (UK, Austria, France, Ireland, Norway, the Netherlands) took part in this process which consisted of three steps.

Systematic literature review

A systematic literature review was conducted to identify original articles examining the validity of B mode or Doppler US in OA of the hand between 1950 and February 2007. The aim was to identify studies that had defined and attempted to measure pathological features of hand OA detectable by US. Pubmed was searched using the terms "ultrasonography and hand and osteoarthritis" and "ultrasound and hand and osteoarthritis". Ovid MEDLINE was also searched using the terms (Osteoarthritis, Hip/or osteoarthritis.mp. or osteoarthritis/or Osteoarthritis, Knee/) and (Hand/or Hand Bones/or hand Joints/or hand.mp./or finger joint/or thumb/or base of

thumb.mb./or metacarpophalangeal joint/or carpal bones/or schapoid bone/or carpal Joints/or stt joint.mp./or finger joint/or pip joint.mp./or dip.m.p.) and (Ultrasonography, Doppler/or Ultrasonography/or Ultrasonography, Doppler, Colour/or Ultrasonography.mp.). The abstracts were reviewed, and articles were excluded if they did not include the use of B mode US in hand OA, were not in English, were case reports, pictorial reviews or review articles, or did not measure pathological features of hand OA.

Iterative internet exercise

The next step was an iterative internet discussion, in order to gain consensus on a draft scoring system developed. Considerations included the joints to be evaluated, domains to be scored, definitions of domains, and proposed scaling systems for each domain. Through this process consensus was obtained on a draft scoring system.

Reliability exercise

A workshop involving 15 experts took place at the Cochin Hospital in Paris, France. This included a multi-observer US reliability exercise to evaluate the draft scoring system. The US reliability exercise involved seven ultrasonographers (DK, RJW, MADA, AP, CSW, HG, HBH) using seven commercially available real time scanners (Technos MPX, Esaote, Genoa, Italy) and a LA435 linear multifrequency transducer of 8–14 MHz. The B mode frequency used was 13 MHz. The power Doppler settings were as follows: frequency of 10 MHz, pulse repetition frequency (PRF) of 750 kHz, and medium wall filter. The colour gain was 113 db and was set at the level at which noise artefacts appeared and then gradually reduced, until only a flow signal, if present, was left. Patients followed in the Department of Rheumatology of Cochin Hospital were invited to participate in this exercise. Seven patients with OA of the hands were scanned by each ultrasonographer to determine inter-reader reliability. Each ultrasonographer rescanned their first patient of the day at the end of the session to assess intra-reader reliability. The US examination imaged 15 joints of the dominant hand; the first carpometacarpal joint, metacarpophalangeal joints 1–5, proximal interphalangeal joints 1–5 and distal interphalangeal joints 2–5. The entire dorsal surface of the joint was imaged in grey scale, and Power Doppler was assessed in the dorsal longitudinal plane.

The inter- and intra-reader reliability was assessed according to κ scores, weighted κ scores (κ (w)),[17] percentage exact agreement (PEA) and percentage close agreement (PCA). Arbitrary qualitative labels have previously been assigned to κ values, whereby a κ of <0.2 is slight, 0.21–0.4 fair, 0.41–0.6 moderate, 0.61–0.8 substantial and >0.81 almost perfect.[18] PEA is the percentage of observations that were given the same score, while percentage close agreement is the number of observations that are either given the same score or ±1. We are not aware of any standard qualitative interpretation of PEA and PCA, so have applied the same cut-offs as for the κ values. Inter-reader reliability was assessed with regards to specific agreement as well, which assesses the agreement specific to each category in the domain. This is less likely to be affected by chance than the PEA. In addition, in assessing intra-reader reliability, the percentage of times the second observation was higher than the first observation was assessed (%m2>m1) in order to identify any potential bias.

The distribution of pathology in the subjects scanned was assessed post hoc by examining the results of the reliability exercise. Pathology was deemed present if four of seven scanners (ie, a majority) scored the pathology as present on the dichotomous scale.

RESULTS
Systematic literature review

Twenty-five articles were identified with the search; the abstracts of these articles were reviewed. Twenty-three articles were excluded; six were not in English, five articles did not involve OA of the hand, three articles were reviews, two were pictorial reviews, six articles utilised neither B mode or Doppler US, and one did not attempt to define or measure pathological features of hand OA. The remaining two articles are presented in table 1.

Iterative internet exercise

Consensus was obtained that 15 joints of the hand would be examined. The first carpometacarpal joint, metacarpophalangeal joints 1–5, proximal interphalangeal joints 1–5, and distal interphalangeal joints 2–5. Domains to be scored reflected domains of activity and damage: synovial hypertrophy and effusion, power Doppler signal and osteophytosis.

Synovial hypertrophy and effusion were considered together as a single domain "synovitis". The OMERACT definitions of synovial hypertrophy and effusion developed for RA were applied.[19] It was agreed that grey scale synovitis would be scored as either present or absent (0–1), and also on a semiquantitative scale of 0–3 analogous to the scoring systems developed in RA, where 0 represented no synovitis, 1 mild synovitis, 2 moderate synovitis and 3 severe synovitis.

Power Doppler signal was defined as a signal within a region of grey scale synovitis. It was decided to assess both dichotomous (present/absent, 0–1) and semiquantitative (0–3) scales.

Osteophytes were defined for the purpose of this exercise as cortical protrusions seen in two planes. Osteophytes were again evaluated using both dichotomous and semiquantitative scales (the latter scored at each joints as absent, mild, moderate or severe on a scale of 0–3).

It was decided not to include erosions, cartilage parameters or joint space narrowing because of concerns about reliable definitions, acquisition, current available US technology and feasibility related to duration of scanning (see also Discussion).

Reliability exercise

Intra- and inter-reader reliability for each domain and each scaling system are presented in tables 2 and 3. The intra-reader κ values (table 3) varied from light to almost perfect depending on the observer, domain and scoring scale. It was generally better from dichotomous scales, with substantive to almost perfect PEA for all observers for each domain. The semiquantitative scales generally demonstrated slightly lower κs and PEA. However, the PCA for the semiquantitative scales was substantive for all observers and all domains. While the intra-reader reliability was quite variable between readers; however, as each reader scanned a different patient, the variability should be interpreted cautiously, as those who scanned subjects with less pathology may be expected to have better reliability. The inter-reader κs (table 2) was fair to moderate, and once again, the dichotomous scales were more reliable than the semiquantitative scales. The PEA ranged from fair to almost perfect, once again depending on domain and scale. For synovitis, the PEA was best towards the normal end of the semiquantitative scale.

Table 1 Papers obtained from a systematic literature review of ultrasonography in osteoarthritis of the hand

First author	Year	Number of subjects	Joint(s) evaluated	US features assessed	Feature defined in text	Quantification of the feature	Inter-reader reliability	Intra-reader reliability	Construct validity	Content validity	Criterion validity
Iagnocco[20]	2005	110	PIP and DIP	Erosions	Yes	Yes (number of erosions)	Yes	–	Yes	–	–
Iagnocco[23]	2000	57	CMC	Effusion	Yes	Yes (depth of articular triangle in mm)	–	–	Yes	–	–

CMC, first carpometocarpal joint; PIP, proximal interphalangeal joint; DIP, distal interphalangeal joint; US, ultrasound.

For the other domains, the results are more variable, for example, agreement on osteophyte scores was better at the extremes of the scale (scores of 0 and 3). There was no consistent bias between or within readers with regards to the second set of observations being higher than the first.

The distribution of pathologies in the subjects scanned is demonstrated in table 4. Osteophytosis was seen in all proximal interphalangeal and distal interphalangeal joints, but there was a wide variation of pathology in other joints.

DISCUSSION

US has many features rendering it potentially valuable in investigating structure in hand OA. This preliminary work established that international experts in the field of US, OA and

outcome measures believe it is worthwhile pursuing this imaging technique as a tool in hand OA. Also, even though US is recognised as a valuable tool in imaging joints in inflammatory diseases, the systematic literature review demonstrated a paucity of information on the validity of US in hand OA.

The development of a preliminary US hand scoring system via an iterative internet exercise allowed experts to come to a consensus as to what US detectable abnormalities in hand OA were both important and feasible domains to be included. These were grey-scale synovitis, power Doppler and osteophytosis. It was felt that current technology would not allow cartilage defects or joint space narrowing to be reliably or meaningfully interpreted, despite these being cardinal pathological features of

Table 2 Intra-reader agreement for the domains and scales used in the reliability exercise

Reader	Reliability	Synovitis (0–1 scale)	Synovitis (0–3 scale)	Doppler (0–1 scale)	Doppler (0–3 scale)	Osteophytes (0–1 scale)	Osteophytes (0–3 scale)
1	PEA	80.0	80.0	80.0	73.3	73.3	66.7
	PCA	–	100.0	–	73.3	–	93.3
	% m2>m1	66.6	66.7	0.0	0.0	100.0	100.0
	κ	0.602	0.679	0.471	0.348	0.375	0.561
	κ (w)	–	0.923	–	0.305	–	0.813
2	PEA	80.0	60.0	86.7	80.0	93.3	80.0
	PCA	–	86.7	–	100.0	–	100.0
	% m2>m1	0.0	16.7	0.0	0.0	100.0	66.7
	κ	0.471	0.474	0.667	0.536	0.842	0.719
	κ (w)	–	0.673	–	0.830	–	0.935
3	PEA	73.3	40.0	80.0	26.7	53.5	40.0
	PCA	–	93.3	–	80.0	–	80.0
	% m2>m1	50.0	77.8	66.7	63.6	71.4	77.8
	κ	0.444	0.172	0.571	0.088	0.087	0.172
	κ (w)	–	0.637	–	0.439	–	0.453
4	PEA	66.7	53.3	60.0	53.3	93.3	80.0
	PCA	–	100.0	–	100.0	–	93.3
	% m2>m1	60.0	57.1	33.3	28.6	0.0	100.0
	κ	0.074	0.356	0.211	0.132	–	0.664
	κ (w)	–	0.753	–	0.276	–	0.763
5	PEA	60.0	53.3	86.7	73.3	93.3	73.3
	PCA	–	86.7	–	93.3	–	100.0
	% m2>m1	16.7	28.6	0.0	25.0	0.0	50.0
	κ	0.224	0.255	0.595	0.250	0.762	0.620
	κ (w)	–	0.545	–	0.637	–	0.900
6	PEA	100.0	100.0	100.0	100.0	100.0	93.3
	PCA	–	100.0	–	100.0	–	100.0
	% m2>m1	–	–	–	–	–	100.0
	κ	1.000	1.000	1.000	1.000	1.000	0.911
	κ (w)	–	1.000	–	1.000	–	0.973
7	PEA	73.3	53.3	73.3	66.7	100.0	53.3
	PCA	–	86.7	–	93.3	–	100.0
	% m2>m1	100.0	85.7	100.0	100.0	–	85.7
	κ	0.444	0.327	0.250	0.085	1.000	0.327
	κ (w)	–	0.369	–	0.277	–	0.656

PEA, percentage exact agreement; PCA, percentage close agreement; κ(w), weighted κ; %m2>m1, percentage second scores higher than first score.

Table 3 Inter-reader agreement for the domains and scales used in the reliability exercise

Seven readers	Synovitis (0–1 scale)	Synovitis (0–3 scale)	Doppler (0–1 scale)	Doppler (0–3 scale)	Osteophytes (0–1 scale)	Osteophytes (0–3 scale)
PEA overall	70.8	47.5	72.7	65.0	83.5	54.2
Specific agreement 0	64.8	65.0	81.0	80.8	64.1	65.2
Specific agreement 1	75.0	32.5	51.8	22.3	89.2	37.1
Specific agreement 2	–	36.8	–	31.7	–	39.1
Specific agreement 3	–	37.7	–	13.6	–	67.2
κ	0.398	0.247	0.327	0.229	0.530	0.378

OA. Given that OARSI guidelines recommend that measures of joint space narrowing are recorded in studies of structural outcomes,[10] such a domain may need to be added to the tool in the future. Erosions were also excluded from the tool, largely due to perceived problems with the definition and reliability. US may be a suitable medium to investigate the relationship between erosive and non-erosive OA, although a single study has found US to be less sensitive to erosions in hand OA than CR.[20] Future development of this tool may need to revisit the issue of whether to include erosions or not.

The reliability exercise demonstrated that the preliminary US scoring tool was reliable. Intra-reader reliability (interpreted with PCA or PEA) for the majority of observers and domains was moderate to almost perfect, being generally higher when the dichotomous scale was used. It is important to acknowledge that variation between observers may be confounded by which subject was used to assess intra-reader reliability. Reliability is likely to be best when less pathology exists (as has been found for the inter-reader reliability). As each observer scanned a different subject to determine their intra-reader reliability, results should be compared with caution.

Inter-reader reliability assessed with PEA was moderate to almost perfect, being over 70% for each domain using a dichotomous scale and being most reliable for the osteophyte domain. Even utilising the semi-quantitative scaling, the lowest PEA was 48% for synovitis, being higher for Doppler and

Table 4 Demonstration of the distribution of pathology by joint. Numbers represent the number of joints with pathology detected at each joint site (the maximum number of joints that could be affected at any site is seven, as seven subjects were examined)

Joint	Grey-scale synovitis (n)	Power Doppler signal (n)	Osteophytes (n)
CMC1	4	3	5
MCP1	2	1	3
MCP2	3	0	3
MCP3	1	0	1
MCP4	1	0	1
MCP5	1	0	2
PIP1	6	2	7
PIP2	6	3	7
PIP3	6	3	7
PIP4	6	2	7
PIP5	6	2	7
DIP2	5	3	7
DIP3	4	2	7
DIP4	5	2	7
DIP5	4	1	7

CMC, carpometacarpal; MCP, metacarpal; PIP, proximal interphalangeal; DIP, distal interphalangeal.

osteophytosis. Perhaps as would be expected, agreement was generally greatest in the absence of pathology. Given there was no formal or extensive standardisation process prior to the exercise, these results are extremely encouraging. Finally, problems with the draft system and future directions were identified.

Despite limiting domains and numbers of joints scanned and imaging only the dorsal surface of the joint, scanning each subject took up to half an hour. This compares poorly with scoring radiographs of hand OA, in which the most time-consuming method has been shown to take an average of less than 4 min.[21] The time taken in this exercise includes the acquisition of images (which is not considered when scoring CRs) and may decrease with increasing observer experience. However, it is likely to be one of the major barriers to a feasible US outcome measure in hand OA. We chose to examine 15 joints for the purpose of this preliminary exercise; however, further consideration of which joints to include in a scoring system, and whether to weight certain joints, is needed. The restriction of scoring systems to limited joints, and weighting of joints according to importance or significance requires further clinical and imaging studies are required to determine the relative significance of each joint or joint combinations.

We did not include erosions, joint space narrowing or cartilage defects in this tool. This was not because these domains were not felt to be of structural or pathological importance in OA, but rather because we chose to focus on domains that were felt to be feasible in a multicentre outcome measure given current technology. Erosions can be difficult to visualise due to overlying osteophytes, and in our experience where the cortical surface is very damaged, it can be difficult to determine where an erosion begins and an osteophyte ends. Furthermore the only study examining the validity of US in detecting erosions in hand OA found US to be less sensitive than CR.[20] While cartilage damage is a cardinal feature of hand OA, there are several features of the US appearance of degenerative cartilage, including thinning, transparency and loss of clarity of the interface.[22] As cartilage changes in OA are a spectrum, an appropriate scaling system would have been complex, time consuming, and of uncertain significance given that visualisation of cartilage is limited in the small joints of the hand, even with flexion of the joint, due to joint structure. US can only image surfaces of joints, the joint space in the central portion can be obscured by osteophytes in OA of the hand, so it was felt that trying to quantify joint space with US would not be feasible.

In the early phase of OA cytokines stimulating osteoblasts are released from the chondrocytes. Thus the osteophytes are early signs of pathology in the cartilage. Given the high resolution of US, small osteophytes not detected on CR may be clinically relevant for early diagnosis of OA.

The definitions we used may need further consideration. For example, we chose to use a composite of synovial hypertrophy and effusion to determine synovial inflammation, recognising that the clinical and pathological relevance of the two are uncertain. It was noted that effusion could occur in the absence of synovial hypertrophy, and a scoring system that grades the features separately may be of interest when time is not constrained. In addition, it was noted that Doppler signal could be seen within the capsule, but external to the hypoechoic areas of synovial hypertrophy. This signal was not scored in this exercise.

The determination of the clinical, pathological and prognostic importance of US detected abnormalities in hand OA, and relative importance of the joints involved was beyond the scope of this exercise. Proof of concept and epidemiological studies need to be undertaken to investigate these issues, and in the future the domains included in US outcome measures for hand OA may need to be revised.

This process has been a preliminary step in developing an US scoring tool for hand OA. Very good PEA for dichotomous scales were demonstrated in this exercise; however, the semiquantitative results suggest that a standardisation process could improve agreement. In addition, preliminary exercises might allow for selecting the most reliable observers. Importantly, this process has demonstrated that an US outcome measure suitable for multicentre trials is feasible and likely to be reliable. In addition it has provided a foundation upon which to further develop this tool.

REFERENCES

1. **Felson DT,** Zhang Y. An update on the epidemiology of knee and hip osteoarthritis with a view to prevention. *Arthritis Rheum* 1998;**41**:1343–55.
2. **Leardini G,** Vaccaro E. Osteoarthritis: socioeconomic problems. *Semin Arthritis Rheum* 2005;**34**(6 Suppl 2):35–7.
3. **March LM,** Bachmeier CJ. Economics of osteoarthritis: a global perspective. *Baillieres Clin Rheumatol* 1997;**11**:817–34.
4. **Zhang Y,** Niu J, Kelly-Hayes M, Chaisson CE, Aliabadi P, Felson DT. Prevalence of symptomatic hand osteoarthritis and its impact on functional status among the elderly: The Framingham Study. *Am J Epidemiol* 2002;**156**:1021–7.
5. **Zhang W,** Doherty M, Leeb BF, Alekseeva L, Arden NK, Bijlsma JW, *et al*. EULAR evidence based recommendations for the management of hand osteoarthritis: report of a Task Force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2007;**66**:377–88.
6. **Zhang W,** Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP, *et al*. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT).[see comment]. *Ann Rheum Dis* 2005;**64**:669–81.
7. **Jordan KM,** Arden NK, Doherty M, Bannwarth B, Bijlsma JWJ, Dieppe P, *et al*. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). *Ann Rheum Dis* 2003;**62**:1145–55.
8. **Dougados M.** Symptomatic slow-acting drugs for osteoarthritis: what are the facts? *Joint Bone Spine* 2006;**73**:606–9.
9. **Dieppe P.** Disease modification in osteoarthritis: are drugs the answer?[comment]. *Arthritis Rheum* 2005;**52**:1956–9.
10. **Maheu E,** Altman RD, Bloch DA, Doherty M, Hochberg M, Mannoni A, *et al*. Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. *Osteoarthritis Cartilage* 2006;**14**:303–22.
11. **Wakefield RJ,** Green MJ, Marzo-Ortega H, Conaghan PG, Gibbon WW, McGonagle D, *et al*. Should oligoarthritis be reclassified? Ultrasound reveals a high prevalence of subclinical disease.[see comment]. *Ann Rheum Dis* 2004;**63**:382–5.
12. **Karim Z,** Wakefield R, Quinn M, Conaghan P, Brown A, Veale D, *et al*. Validation and reproducibility of ultrasonography in the detection of synovitis in the knee: a comparison with arthroscopy and clinical examination. *Arthritis Rheum* 2004;**50**:387–94.
13. **Szkudlarek M,** Narvestad E, Klarlund M, Court-Payen M, Thomsen HS, Ostergaard M. Ultrasonography of the metatarsophalangeal joints in rheumatoid arthritis: comparison with magnetic resonance imaging, conventional radiography, and clinical examination. *Arthritis Rheum* 2004;**50**:2103–12.
14. **Lopez-Ben R,** Bernreuter WK, Moreland LW, Alarcon GS. Ultrasound detection of bone erosions in rheumatoid arthritis: a comparison to routine radiographs of the hands and feet. *Skeletal Radiol* 2004;**33**:80–4.
15. **Wakefield RJ,** Gibbon WW, Conaghan PG, O'Connor P, McGonagle D, Pease C, *et al*. The value of sonography in the detection of bone erosions in patients with rheumatoid arthritis: a comparison with conventional radiography.[see comment]. *Arthritis Rheum* 2000;**43**:2762–70.
16. **Szkudlarek M,** Court-Payen M, Strandberg C, Klarlund M, Klausen T, Ostergaard M. Power Doppler ultrasonography for assessment of synovitis in the metacarpophalangeal joints of patients with rheumatoid arthritis: a comparison with dynamic magnetic resonance imaging. *Arthritis Rheum* 2001;**44**:2018–23.
17. **Fleiss JL,** Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973;**33**:613–19.
18. **Landis RJ,** Koch GG. The measurement of Observer Agreement for Categorical Data. *Biometrics* 1977;**33**:159–74.
19. **Wakefield RJ,** Balint PV, Szkudlarek M, Filippucci E, Backhaus M, D'Agostino MA, *et al*. Musculoskeletal ultrasound including definitions for ultrasonographic pathology. *J Rheumatol* 2005;**32**:2485–7.
20. **Iagnocco A,** Filippucci E, Ossandon A, Ciapetti A, Salaffi F, Basili S, *et al*. High resolution ultrasonography in detection of bone erosions in patients with hand osteoarthritis. *J Rheumatol* 2005;**32**:2381–3.
21. **Maheu E,** Cadet C, Gueneugues S, Ravaud P, Dougados M. Reproducibility and sensitivity to change of four scoring methods for the radiological assessment of osteoarthritis of the hand. *Ann Rheum Dis* 2007;**66**:464–9.
22. **Grassi W,** Lamanna G, Farina A, Cervini C. Sonographic imaging of normal and osteoarthritic cartilage. *Semin Arthritis Rheum* 1999;**28**:398–403.
23. **Iagnocco A,** Coari G. Usefulness of high resolution US in the evaluation of effusion in osteoarthritic first carpometacarpal joint. *Scand Journal of Rheumatol* 2000;**29**:170–3.