CONCISE REPORT

Assessing the intra- and inter-reader reliability of dynamic ultrasound images in power Doppler ultrasonography


See linked article, p 1590

Objective: To assess the intra-reader and inter-reader reliabilities of interpreting ultrasonography by several experts using video clips.

Method: 99 video clips of healthy and rheumatic joints were recorded and delivered to 17 physician sonographers in two rounds. The intra-reader and inter-reader reliabilities of interpreting the ultrasound results were calculated using a dichotomous system (normal/abnormal) and a graded semiquantitative scoring system.

Results: The video reading method worked well. 70% of the readers could classify at least 70% of the cases correctly as normal or abnormal. The distribution of readers answering correctly was wide. The most difficult joints to assess were the elbow, wrist, metacarpophalangeal (MCP) and knee joints. The intra-reader and inter-reader agreements on interpreting dynamic ultrasound images as normal or abnormal, as well as detecting and scoring a Doppler signal were moderate to good (k = 0.52–0.82).

Conclusions: Dynamic image assessment (video clips) can be used as an alternative method in ultrasonography reliability studies. The intra-reader and inter-reader reliabilities of ultrasonography in dynamic image reading are acceptable, but more definitions and training are needed to improve sonographic reproducibility.

Grey scale and Doppler ultrasound imaging are useful methods for locating soft-tissue lesions of synovial structures such as joints, tendons and bursae, as well as bone erosion, and determining any inflammatory changes. Ultrasonography has the reputation of being a very operator-dependent technique.

Rheumatologists have put a lot of effort into assessing intra-reader and inter-reader/observer reliability in interpreting still images, as well as image acquisition in ultrasound depiction. In most studies, ultrasonographic inter-observer reliability has been tested between two observers. Intra-reader and inter-reader agreement in Doppler imaging has not been tested with dynamic image reading (video clips).

The aim of this paper was to test video-clip reading as a means of assessing ultrasound results and to evaluate the intra-reader and inter-reader reliabilities of assessing dynamic images of healthy and rheumatic joints for normal and pathological states, as well as detecting and scoring a Doppler signal.

MATERIALS AND METHODS

An Esaote Technos ultrasound system (Esaote Biomedica, Genova, Italy) was used. The system was equipped with two linear probes: LA424 (frequency range 8–14 MHz) and LA523 (frequency range 5–10 MHz). The first probe was used in hand and foot joints, and the second in elbow, shoulder and knee joints.

Ultrasound scanning, video recording and a percutaneous synovial biopsy of the site scanned were carried out by JM K on 41 patients with monoarthritis or polyarthritis in 41 synovial sites: 22 knee, 7 wrist, 3 tibiotalar, 2 metatarsophalangeal (MTP), 1 glenohumeral, 1 metacarpophalangeal (MCP), and 1 elbow joint as well as 2 subdeltoid bursae, 1 tibialis posterior and 1 peroneus tendon sheath. The clinical characteristics, scanning procedures, biopsy methods and histopathological evaluation have been reported in Koski et al.[12] All the joints, except for one, were abnormal in histology. An abnormal sonography result was obtained in 98% of the patients and the power Doppler was positive in 77% of the cases.[3] Furthermore, 58 video clips of joints of healthy people were recorded. These people were asymptomatic volunteers with no pre-existing joint trauma or disease, and their clinical status was normal. We were not able to collect as many arthritic cases as normal volunteers. However, we decided to include all normal cases, as this would increase the reliability of statistical analysis. In all 40 of the volunteers were women and 18 men. Their mean age was 40 years (18–65 years). In total, 7 MCP, 11 wrist, 7 elbow, 8 shoulder, 14 knee, 5 tibiotalar and 6 metatarsophalangeal joints were scanned and recorded by UH. The probe positions and the areas recorded corresponded exactly between the healthy and patient groups when standard scans by EULAR[13] were used.

During the video recording of the region of interest, the probe was left immobile to avoid motion artefacts in Doppler imaging. The digital video camera connected to the Ultra Sound (US) equipment was a Sony DCR-TRV 900E (Sony Corporation, Tokyo, Japan). UH and JMK edited a CD ROM, mixing normal and patient clips randomly. Thus, the CD ROM included 99 video clips lasting, on the average, 13.1 (SD 4.4) s in the healthy group and 19.7 (5.7) s in the patient group (p<0.01). A copy of the CD ROM was delivered to 17 physician sonographers in Europe (round one). A second CD ROM with the same video clips but a different randomisation was sent to the same readers after 3–4 months (round two). In the meantime, they were not allowed to watch the first CD ROM. The readers did not know whether the clips were from normal or pathological. They only knew which joint site was involved and the orientation of the transducer. First they gave an anonymous answer to a question on a preformatted documentation sheet: “Do you see a Doppler signal?” A semiquantitative subjective grading from 0 to 3 was used: 0 signified no detectable Doppler signal inside the synovium (only) of the joint bursa or tenosynovium; 1, mild but clear; 2, moderate and 3, substantial increase in Doppler signal. Secondly, they answered yes or no to the question: “Is the case from a normal person or a patient with an inflammatory...
joint disease?". Here they were allowed to evaluate the grey scale changes of bony surfaces, effusion, synovial proliferation and the Doppler signal.

**Statistical analysis**

Statistical analyses were carried out using SPSS V.13 software. An independent samples t test was used to determine the difference between the durations of the video clips. Spearman’s ρ correlation analyses between variables were tested for two-tailed probability values. Values of p<0.05 were considered significant. Intra-reader and inter-reader agreements were assessed by calculating a κ coefficient between the readers.14 15 κ coefficients were classified as follows: <0, poor; 0.00–0.20, slight; 0.21–0.40, fair; 0.41–0.60, moderate; 0.61–0.80, substantial; and 0.81–1.00, almost perfect agreement.

**RESULTS**

In all, 70% of the readers could classify at least 70% of the videos correctly as belonging to the healthy control group or the patient group. The distribution of readers answering correctly was wide (fig 1). Intra-reader agreement was found to be good to excellent and inter-reader agreement was found to be moderate to good (table 1). The elbow, wrist, MCP and knee joints were the most difficult ones to assess (fig 2).

Statistically significant correlation between the semiquantitative evaluation of the strength of the Doppler signal and the histological score of synovitis in any of the readers was found to be nil: mean correlation 0.17 (range 0.11–0.24) in the first round and 0.17 (range 0.01–0.23) in the second round.

**DISCUSSION**

The interpretation of the US videos was clearly reader-dependent. Two readers classified the videos correctly as normal or abnormal in about 90% of cases on both rounds. On the other hand, almost one third of the readers could do this only in about 60% of cases. Intra-reader agreement was good to excellent, whereas the inter-reader agreement was moderate to good. Because of the small number of cases in the subgroups, reliability in different joints was not calculated. The results were quite similar to those reported in earlier studies with several readers or observers.9–11 The video reading method seemed to work well. More definitions are needed to raise the level of the results. Defined calibration images could also improve the inter-reader variability. Like the principal sonographer JMK,12 the 17 video readers did not find significant statistical correlations between the severity of histological synovitis and Doppler signal.

The primary goal of this study was to examine power Doppler ultrasound imaging. However, the Doppler signal is only a part of the ultrasound image and thus grey scale ultrasound also had to be taken into account in evaluating the images.

The best way to test operator dependence between several observers is for each examiner to perform the scanning blindly (the image acquisition). In the present study, this arrangement was not possible. We used video clips instead of still images, because Doppler imaging is a dynamic method and a video gives a better impression of the live situation. The advantages of the video reading method are: (1) compared with image acquisition, sample size is large; (2) readers are fully blinded to whether the joint is from a patient or a healthy person; (3) the second round of reading can be easily organized; and (4) a copy of the CD ROM can be delivered to several countries and readers. Furthermore, the length of a video clip should be the same in normal and abnormal cases. We could not achieve this in the present study.

In conclusion, dynamic image reading (video clips) is an alternative method for studying reliability in sonography. The intra-reader and inter-reader reliabilities of interpreting dynamic ultrasound images for classifying cases as normal or abnormal, as well as detecting and scoring Doppler signals in the synovium, are moderate to good, but more definitions and training are needed.
Correspondence to: J Koski, Mikkeli Central Hospital, Porrassalmienkatu 35–37, Fin-50100 Mikkeli, Finland; f.koski@fimnet.fi

Accepted 15 May 2006

REFERENCES


Assessing the intra- and inter-reader reliability of dynamic ultrasound images in power Doppler ultrasonography


Ann Rheum Dis 2006 65: 1658-1660 originally published online May 25, 2006
doi: 10.1136/ard.2005.051250

Updated information and services can be found at: http://ard.bmj.com/content/65/12/1658

These include:

References
This article cites 15 articles, 5 of which you can access for free at: http://ard.bmj.com/content/65/12/1658#BIBL

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections

- Clinical diagnostic tests (1279)
- Radiology (1112)
- Radiology (diagnostics) (750)

Notes

To request permissions go to: http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to: http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to: http://group.bmj.com/subscribe/