Intraobserver repeatability and interobserver reproducibility in musculoskeletal ultrasound imaging measurements

P.V. Balint, R.D. Sturrock

Peter V. Balint, MD; Roger D. Sturrock MD, FRCP

Centre for Rheumatic Diseases, University Department of Medicine, Glasgow Royal Infirmary

The Centre for Rheumatic Diseases is supported by Arthritis Research Campaign.

Please address correspondence and reprint requests to: Peter V. Balint, Centre for Rheumatic Diseases, University Department of Medicine, Glasgow Royal Infirmary, QEB, 10 Alexandra Parade, Glasgow, G31 2ER, Scotland, UK
E-mail: pb58v@clinmed.gla.ac.uk

Received on November 13, 2000; accepted on January 3, 2001.

Key words: Reproducibility of results, observer variation, ultrasonography, diagnostic imaging, musculoskeletal system, hip joint, hip, ligaments, phantoms, imaging.

ABSTRACT

Objective. *To assess the repeatability and reproducibility of ultrasonographic measurements at the anterior surface of the femoral neck and iliofemoral ligament and on a human tissue-mimicking phantom.*

Methods. *Two independent investigators studied 22 consecutive hips. One investigator had previous experience in musculoskeletal ultrasonography (US). The other investigator had undergone a short course in hip sonography (only 3 hours). Both investigators were blinded to their own and each other's results. On the phantom both observers had taken 10 vertical measurements at 6 cm deep where two objects were placed at 2 cm from each other. Calculation of measurement errors, percent errors and the Bland-Altman graphic technique were used for analysis of data.*

Results. *After 132 examinations the first investigator's within-subject standard deviation was 0.4 mm. The intraobserver error was 4.75%. The second investigator's within-subject standard deviation was 0.6 mm and his intraobserver error was 7.00%. The interobserver error was 10.91%. After 20 phantom examinations the first investigators's intraobserver error was 1.11% and the second investigator's intraobserver error was 1.47%.*

Conclusion. *An inexperienced musculoskeletal sonographer can achieve an acceptable performance if given appropriate training.*

Introduction

It is a common view that one of the major disadvantages of musculoskeletal ultrasound (US) is operator-dependency (1-8). In musculoskeletal US imaging the images generated are mainly qualitative and agreement has to be reached by different observers as to the presence or absence of pathological signs or disease. If quantitative measurements are required, then intra- and interobserver errors become more important. We have therefore determined the magnitude of inter- and intraobserver errors using US imaging for the measurement of the distance between the iliofemoral ligament and the femoral neck in 22 hip joints from an unselected group of normal controls and patients with inflammatory joint disease. Individuals with a history of previous hip surgery were excluded from the study. The hip joint was selected for the following reasons:

1. The hip is a deep joint and not easy to palpate. Hip joint effusions are not easily detected by clinical examination. However the iliofemoral ligament and the neck of the femur are easily identified on US imaging.
2. There is an extensive literature describing the US appearances of the anterior hip joint recess in health and disease, but only one non-blind study calculated intra- and interobserver errors (9).

In addition, an assessment of intraobserver error was measured using a phantom containing two wires at 4 and 6 cm deep from the surface and measuring the vertical depth between these wires.

Materials and methods

Two independent investigators studied 22 hips. One investigator (PVB) had previous experience in musculoskeletal ultrasonography (US). The other investigator (RDS) had undergone a short course in hip sonography (only 3 hours). Each hip was studied with an ATL (Advanced Technology Laboratories, Bothell, Washington,USA) HDI (High Definition Imaging) 3000 ultrasound machine with a linear (L) 7-4 MHz 38-mm footprint probe and musculoskeletal software. US imaging was performed in the oblique sagittal plane from an anterior approach with the subject in a supine position with the straight leg in slight external rotation (Fig 1). Normal anatomical reference landmarks were established (head of the femur, neck of the femur, iliofemoral ligament). The femoral neck-iliofemoral ligament distance was measured in triplicate in quick succession (Fig 2). Before each measurement a new image was generated and the measurements taken.

Using the phantom (Gammex RMI 403GS, Middleton,WI, USA) both observers took 10 vertical measurements at a depth of 6 cm where two wires were placed at 2 cm from each other

Fig. 1. Picture showing the standard position of the probe for US imaging of the hip.

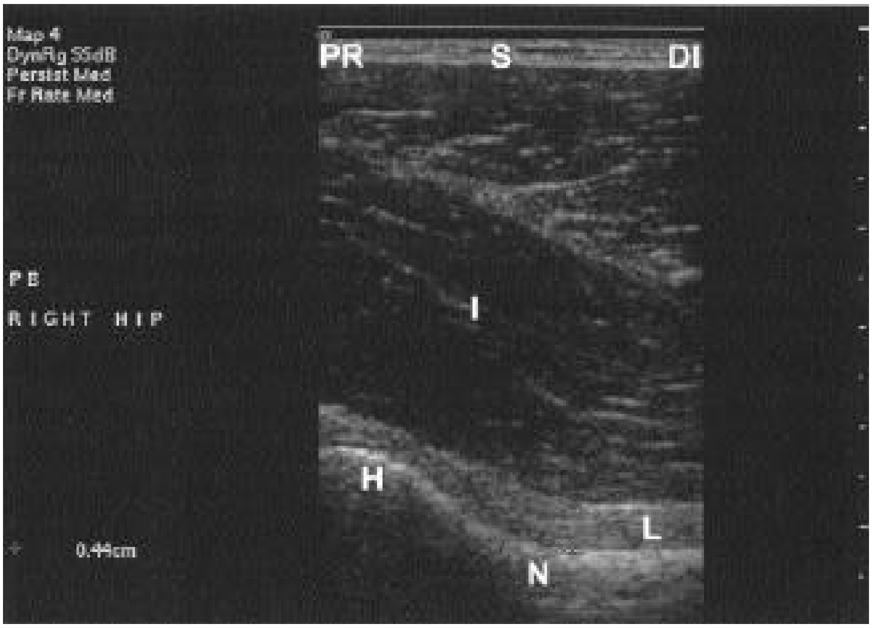

Fig. 2. Hip ultrasound image. The crosses indicate the position of the iliofemoral ligament and the femoral neck. **H**: femoral head, **N**: femoral neck, **L**: iliofemoral ligament, **I**: iliopsoas muscle, **S**: skin, **PR**: proximal end of probe, **DI**: distal end of probe

Fig. 3. Phantom. The crosses mark the position of the wire markers and the dotted line is the distance between the two markers.

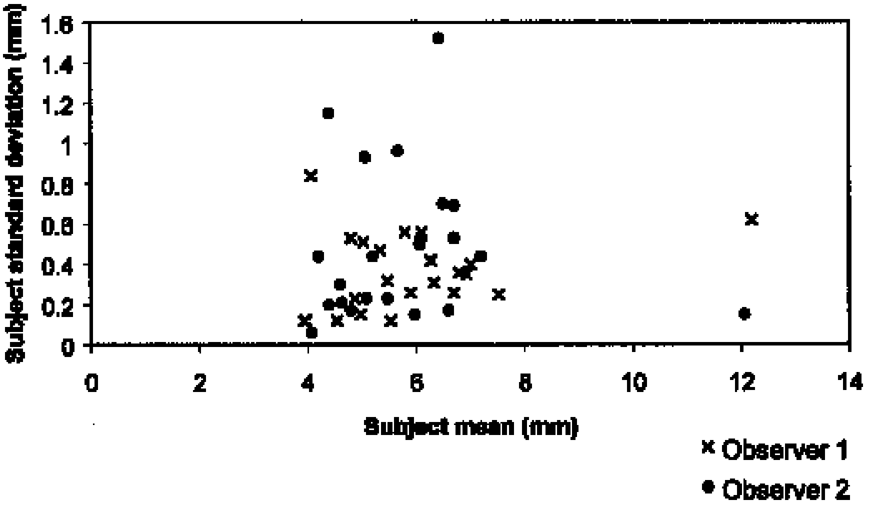

Fig. 4. A plot of the standard deviation of the observers measurement of iliofemoral thickness against the mean of the triplicate values for each hip.

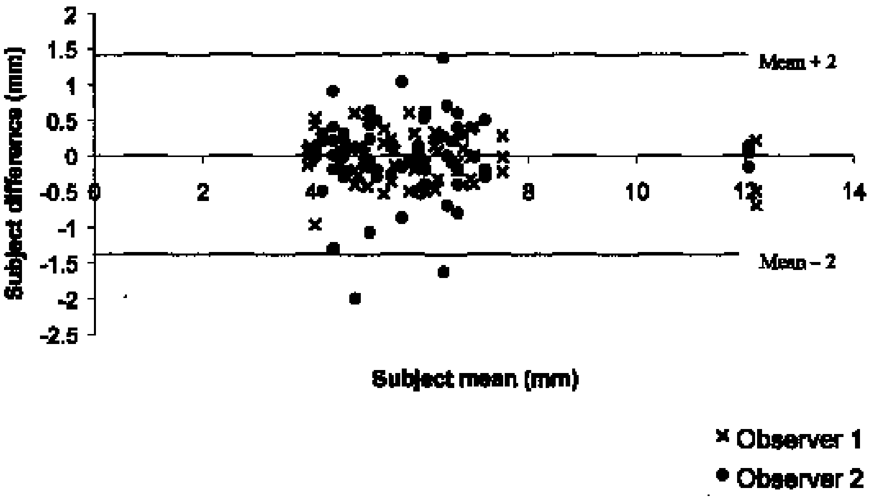

Fig. 5. Intraobserver comparisons: This shows the difference between the triplicate measurements of each hip plotted against the mean value obtained at each hip examination for the two observers.

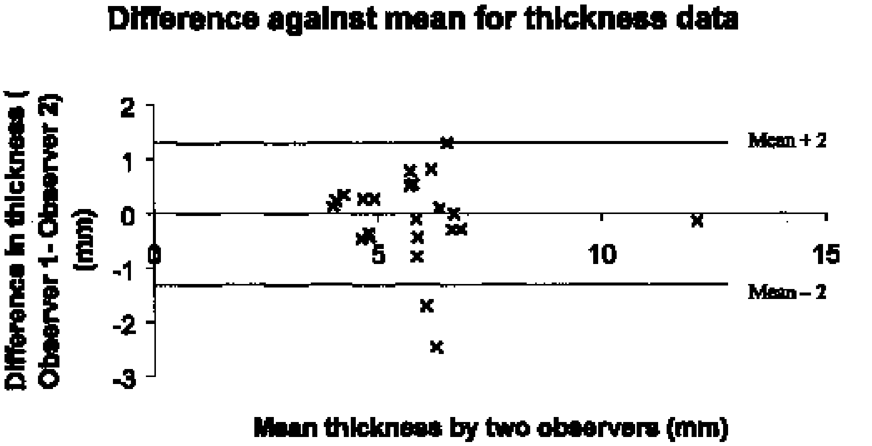

Fig. 6. Interobserver comparison of iliofemoral thickness. A plot of the difference between the observers mean thickness measurements against the overall mean thickness for the two observers (Bland-Altman plot).

(Fig.3). Again both investigators were blinded to their own and each other's results. Each observer's measurement errors were calculated with within-subject standard deviations. A plot diagram was used to show that the observer's standard deviations were unrelated to the magnitude of the measurement (10). Correlation coefficients were used to assess the linear relation of the two sets of mean measurements between the two observers. We used the Bland-Altman graphic technique to assess the agreement between two observers (11). Phantom measurements were analysed as a percentage of deviation from the known true value.

Results

A total of 152 images were recorded and every image was readable.

To obtain the common within-subject standard deviation (s_w) we averaged the variances and the squares of standard deviations. The first investigator's (s_w) was 0.4 mm. A plot diagram was used to prove that the subject's standard deviations are unrelated to the magnitude of the measurement of iliofemoral ligament (Fig. 4). The difference between measurements for the same subject is expected to be less than 2.77 s_w for 95% of pair measurements. Figure 5 shows the differences against their means. Intraobserver error was also expressed with discrepancies from their means in percentages. The difference of the higher value and the lower value divided by the lower value multiplied by 100 gives the individual percent error. The mean of these individual values over all 22 cases gives the final result. In this case the intraobserver error was 4.75%.

The second investigator's (s_w) was 0.6 mm. Figure 4 shows on a plot diagram that the subject's standard deviations are again unrelated to the magnitude of the measurement.

Figure 5 shows the differences against their means. Intraobserver error was also expressed in percentage of deviation from the means. In this case the intraobserver error was 7.00%. For calculating the interobserver error both investigators' mean values were used. The correlation coefficient was also calculated and there was a relation between the values (r = 0.89). The agreement between the two observers was measured with the Bland-Altman method (Fig. 6). The interobserver error was 10.91%. Both investigators were tested blindly on the phantom object and the first investigator's average depth was 19.78 mm, S.D. 0.15 mm. Percent error was 1.11%. The second investigator's average depth was 19.71 mm, S.D. 0.17 mm. The percent error was 1.47%.

Discussion

In this study every US image obtained was of acceptable quality. With well-defined anatomical landmarks and with pre-determined criteria the interobserver variation between the two observers was acceptable. However, for US of the hip measurements were taken in the sagittal plane only, as this is the standard approach for hip US. Most US imaging is performed in two different planes, which might lead to greater interobserver errors at the same depth. None of the patients studied weighed more than 90 kg and it is well known that US imaging of the hip is more unreliable in obese subjects and therefore more likely to increase the possibility of interobserver variation. The positioning of US probes is critical in obtaining an interpretable US image and a slight alteration in the angle of the probe in relation to the skin surface or a variation in the amount of gel used can greatly distort the image obtained and increase the occurrence of artefacts. Musculoskeletal US is now becoming a tool increasingly used by rheumatologists (12-15) most of whom have had no formal training in imaging techniques. This study demonstrates that a rheumatologist with experience of US imaging can train a novice within a relatively short space of time to produce acceptable images of the hip and with relatively small interobserver variation. Further studies will be required to assess whether this is possible for more complex joints such as the shoulder.

Acknowledgements

We gratefully acknowledge George Corner, Daniel Feszty, Alastair Gracie, John Lockhart and Agnes Ory for their technical assistance.

References

1. MIDDLETON WD, REINUS WR, MELSON GL, TOTTY WG, MURPHY WA: Pitfalls of rotator cuff sonography. *Am J Roentgenol* 1986; 146: 555-60.
2. MACK LA, NYBERG DA, MATSEN FA III: Sonographic evaluation of the rotator cuff. *Radiol Clin North Am* 1988; 26: 161-77.
3. HALLFM: Sonography of the shoulder. *Radiology* 1989; 173: 310.
4. BRANDT TD, CARDONE BW, GRANT TH, POST M,WEISSCA: Rotator cuff sonography: A reassessment. *Radiology* 1989; 173: 323-7.
5. BURK DL JR, KARASICK D, KURTZ AB *et al.*: Rotator cuff tears: Prospective comparison of MR imaging with arthrography, sonography, and surgery. *Am J Roentgenol* 1989; 153: 87-92.
6. SOBLE MG, KAYEAD, GUAY RC: Rotator cuff tear: Clinical experience with sonographic detection. *Radiology* 1989; 173: 319-21.
7. HODLER J, TERRIER B, VON SCHULTHESS GK,FUCHS WA: MRI and sonography of the shoulder. *Clin Radiol* 1991; 43: 323-7.
8. JACOBSON JA, VAN HOLSBEECK MT: Musculoskeletal ultrasonography. *Orthop Clin North Am* 1998; 29: 135-67.
9. KOSKI JM,ANTILLA PJ, ISOMAKI HA: Ultrasonography of the adult hip joint. *Scand J Rheumatol* 1989; 18: 113-7.
10. BLAND JM,ALTMANDG: Measurement error. *BMJ* 1996; 313: 744.
11. BLAND JM,ALTMAN DG: Statistical methods for assessing agreement between 2 methods of clinical measurement. *Lancet* 1986;1:307-10.
12. MANGER B, KALDEN JR: Joint and connective tissue ultrasonography - A rheumatologic bedside procedure ? *Arthritis Rheum* 1995; 38: 736-42.
13. BALINT P, STURROCK RD: Musculoskeletal ultrasound imaging:A new diagnostic tool for the rheumatologist ? *Br J Rheumatol* 1997; 36: 1141-2.
14. GRASSI W, CERVINI C: Ultrasonography in rheumatology: An evolving technique. *Ann Rheum Dis* 1998; 57: 268-71.
15. WAKEFIELD RJ, GIBBON WW, EMERY P: The current status of ultrasonography in rheumatology. *Rheumatology* 1999; 38 :195-8.