Digital image analysis protocol for determining the radiocarpal joint space in the rheumatoid arthritic wrist

Wilkinson SJ[1], Naylor A[2,3], Goh KL[2,3]*

[1] The School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, Scotland

[2] School of Mechanical and Systems Engineering, Newcastle University, Newcastle upon Tyne, England NE1 7RU, UK

[3] NU International Singapore, 172A Ang Mo Kio Avenue 8, #05-01, SIT Building @ Nanyang Polytechnic, Singapore 567739

* Corresponding author: email kheng-lim.goh@ncl.ac.uk

Tel +65 6908 6073

Keywords: Dual-energy x-ray absorptiometry; rheumatoid arthritis; radiocarpal joint

Abstract

This paper describes a simple protocol for measuring the joint space of the rheumatoid arthritic (RA) wrist from projection radiographs. The protocol is implemented using a computer algorithm based upon the Interactive Data Language platform. The computerized algorithm features a user-friendly graphical interface to aid the operator to measure joint space parameters, namely distance and area, of the wrist vertebral morphometry at the radiocarpal region. Dual-energy X-ray absorptiometry (DXA) radiograph of a standard hand and wrist phantom was evaluated using the measurement protocol to determine the accuracy and precision of the protocol. The accuracy, parameterised by the systematic error, returned a mean of 5.20 % for distance and is equal to 3.49 % for area measurement. The precision of the measurement protocol, parameterised by the coefficient of variation (CV), for distance returned a mean of 1.96 %; the CV for area measurement equals 2.1 %. Three observers participated to investigate the repeatability (intra-observer) and reproducibility (inter-observer) of the measurement protocol, parameterised by the CV, using DXA radiographs from a healthy volunteer and a RA patient. The inter-observer repeatability for distance measurement for the respective observers returned mean values of 10.9 %, 7.7 % and 11.4 % for the healthy wrist. However, the results revealed improved repeatability for the RA wrist; the CV for the respective observers returned mean values of 7.7 % 7.1 % and 10.0 %. The inter-observer repeatability for area measurement for the respective observers returned mean values of 10.2 %, 7.1 % and 10.1 % for the healthy wrist. However, the results revealed improved repeatability (in two out of the three observers) for the RA wrist; the CV for the respective observers returned mean values of 6.8 % 6.5 % and 10.8 %. Student’s t-test analysis of the intra-observer repeatability revealed that the measurements of distance and area

were generally not intra-observer sensitive. On the other hand, student’s t-test analysis of the inter-observer reproducibility revealed that half of the distance measurements were inter-observer sensitive; whereas the remaining were not. Similar findings were obtained for area measurements. Overall the results reveal that the variabilities in accuracy and precision tests and the repeatability and reproducibility tests were typically 10% or less. These findings, in addition to the versatility and simplicity of the digital image analysis protocol, lend to the potential of using the protocol to complement the acquisition of bone mineral density data derived from DXA for diagnosing the progression of RA in patients.

1 Introduction

Rheumatoid Arthritis (RA) is a chronic progressive disease resulting in joint inflammation, and consequently extreme discomfort and pain. RA typically starts in the membrane surrounding the joint (the synovium), which then thickens and fills the joint space (JS). The joint function deteriorates further with age [1]; RA is more common amongst women over 50 years of age than any other demographic [2]. Radiography plays an important role in the diagnosis of RA as the features associated with the pathology of the diseases can be visually assessed [3]. Protocols for radiologic assessment of RA can be broadly classified into: (1) scoring system; and (2) quantitative radiography. The “gold standard” has been the scoring systems initially proposed by Kellgren and Lawrence [3], but more recently implemented by other authors [4-7]. However, a major limitation of any scoring system is that it is qualitative and hence inherently subjective [4]. It is for this reason that quantitative radiography has been proposed as an alternative since it aims to measure the distribution and size of each radiographic feature accurately and precisely [8]. Quantitative radiography to assess radiographic structural changes, such as JS narrowing at the peripheral JS of RA patients, is intended to measure cartilage loss from erosion effects [4]. Typical measurement parameters for quantitative radiography are distance [4] and area [9] and these parameters may be evaluated directly from the radiograph [10] or from computerized image analysis of digitized x-ray films [9]. However, the accuracy and precision of the diagnosis depends very much on the measurement protocol [4, 9, 11].

Additionally, poor contrast and resolution in conventional x-ray systems have hindered their usefulness in quantitative radiographic assessment despite their low-cost, ease of interpretation and ability to provide a permanent record which can be assessed at any stage of the disease [8]. Alternative systems have been proposed and developed. Buckland-Wright [12] developed a microfocal radiographic system capable of producing high-definition images, resolution and have applied the system to both knee and hand [11]. Harvey et al. [13] employed a second generation dual-energy x-ray absorptiometric (DXA) scanner to achieve high contrast in radiographs of bone and used interactive computerized image analysis technique for accurate delineation of bony margin in the hip joint for distance measurement. DXA belongs to a family of multi-energy imaging techniques that are designed to acquire radiographs containing energy-independent information free from artifacts caused by x-ray attenuation as seen in conventional x-ray radiography [14-16].

More important, DXA scanners are used for measuring bone mineral density (BMD) to assist in the diagnosis of bone deterioration and loss [17-20] that could be linked to RA [13, 18]. To this end, this paper proposes a computerised image analysis method for the assessment of JS in the wrist of RA patients using digital radiographs acquired based on existing imaging protocols. A computer algorithm (known as 'WRISTJS') was developed to implement the digital image analysis protocol. To measure the radiocarpal JS of the wrist from radiographic images, the proposed measurement protocol involved evaluating distances and areas within the JS. The distance parameter measures the separation between joint margins, and provides a measure of JS narrowing at specific sites along the radius and carpal bones of the wrist. The area parameter measures compartmentalised areas of the JS of the respective radius and carpal bones of the wrist to facilitate overall assessment of the degree of JS narrowing at these sites. The protocol for the JS measurement of joint space is driven by the need to avoid excessive and complicated procedures. Here, each stage of the protocol is executed independently; this reduces the potential effects of sequential systematic error. It is envisaged that such a method to acquire data related to distances and areas with the joint space of the RA wrist could be used to complement the BMD results obtained from the same DXA images for diagnosing the progression of the disease.

2 Materials and methods

2.1 Reference locations

A fundamental approach in the quantitative radiographic measurement of the wrist JS is the use of reference locations to achieve precision [11, 21] or the use of landmarks identified by active shape models [21]. In this study the focus is on manual identification of reference locations using anatomical landmarks (ALs).

Ideally, the reference locations must be radiographically visible at any stage during the period of study of the disease. These reference locations (or landmarks) play an important role in ensuring that: (1) the same observer will obtain consistent repeated measurements performed on the same region; (2) different observers will obtain measurements from the same image with minimal variability; (3) temporal changes in any radiographic feature are detectable. In this study, six ALs along the bony margin in the radiocarpal JS have been identified for the measurement protocol. These ALs are indicated in the schematics in Figure 1. The observers were instructed on the arguments underpinning these ALs; this would help them make informed decision when identifying and marking them digitally during the image analysis process.

Two ALs, namely #1 and #2, along the leading bony edge of the radius medial and lateral ends have been designated as the radius “tangential” locations (Figure A). As the name suggests, they have been defined such that if a straight line is drawn passing through them, the line should just rest on the bony margin. Another two ALs (#3, #4) along the bony margins of the scaphoid and lunate have been designated as the carpal "tangential" points (Figure 1 A)—if a straight line is drawn passing through them, the line should just rest on the bony margin of both scaphoid and lunate.

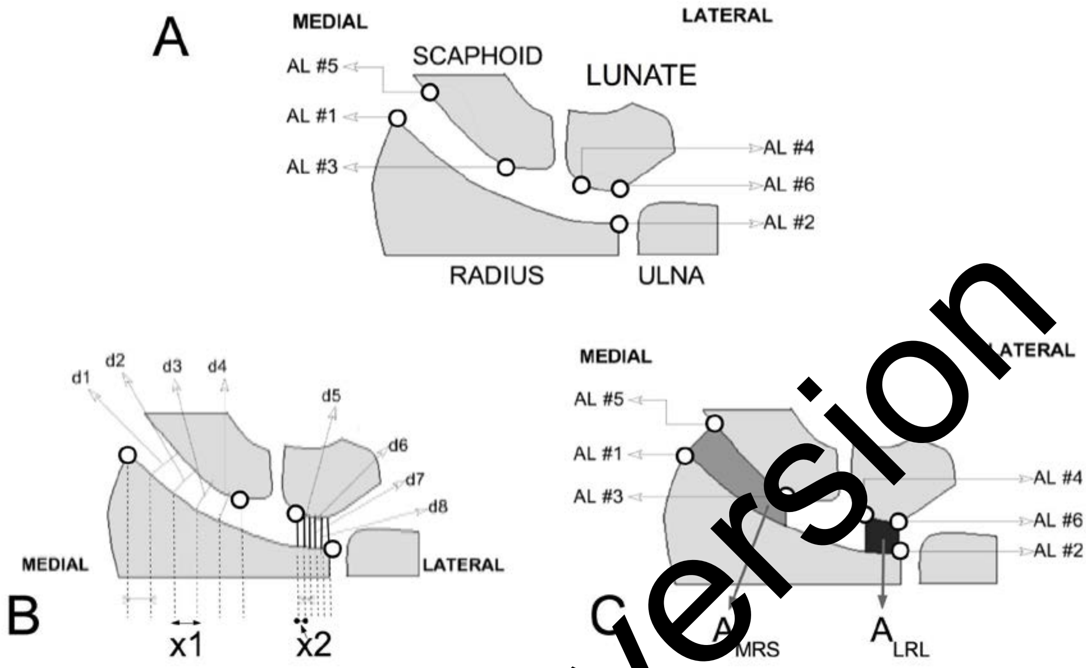

Figure 1. Schematics of the digital image analysis protocol for rheumatoid arthritic assessment. (A) Anatomical locations (ALs) at the radiocarpal joint space. AL #1 and #2 are identified by the points of contact of a straight line with the two leading edges of the radius bony margin. The same applies with respect to the scaphoid and lunate bony margins for both AL #3 and #4. AL #5 is identified on the scaphoid bony margin by the "intersection" of the shortest straight line joining AL #1 and the bony margin; the same rule is applied to the lunate for AL #6 in relation to AL #2. (B) Distance measurement. d1-d4 and d5-d8 refer to the joint space distances between the radius and scaphoid and the radius and lunate respectively; x1 denotes distance between equally spaced lines within AL #1 and #3; x2 denotes equally spaced lines between AL #4 and #2. (C) Area measurement of the medial radioscaphoid (MR) region, AMRS, and the lateral radiolunate (LR) region, ALRL.

The last two ALs, i.e. #5 (along the medial scaphoid) and #6 (along the lateral lunate carpal) have been designated as the carpal "end" location (Figure 1 A). AL #5 has been defined on the scaphoid bony margin by the intersection point of the carpal bone with the radius leading edge. Although AL #6 could be defined by the intersection point of a straight line (drawn perpendicular to the bony margin where AL #2 is located) with the lunate bony margin, in some cases it was more appropriate to define AL #6 as an intersection point of the carpal bone with the radius leading edge. This is for when it was more feasible to detect the overlapping projected images of the radius and carpal bones.

After the ALs were identified and registered into WRISTJS, the radiocarpal JS would be compartmentalized into the radioscaphoid and radiolunate region to facilitate independent assessment of RA. The anatomical locations (#1, #2, #4 and #6) on the carpal and radius bone defined the horizontal bounds of each compartment; the curves outlined by the bony margin define the vertical bounds of the compartments.

2.2 Distance and area measurement

As mentioned in section 1, the two quantities used for JS measurement are distance and area. For the purpose of illustrating how WRISTJS executes the distance measurement, consider the radioscaphoid compartment (Figure 1). Here, WRISTJS has been programmed to divide the horizontal (with respect to the image axes) spatial separation between the anatomical location #1 on the radius bone and the anatomical location #3 on the scaphoid bony margin into five equal segments. From these segments, four equally spaced (at distance of Δx in the horizontal direction) points along the curve fitted to the radius would be identified. Rudimentary trigonometric methods were used to determine the distances (d1, d2, d3 and d4) between the radius and the scaphoid (Figure 1 B).

The arguments used to implement the area calculations were more straight-forward: the areas of interest were the regions enclosed by the radiolunate and radiocarpal compartments. The areas of regions A_{MRS} and A_{LRL} were taken as the sum the areas constrained by d1 to d4 between the radius and scaphoid; and d5 to d8 between the radius and lunate respectively.

2.3 Measurement protocol

The protocol for measuring the wrist JS involves the following steps: (1) image processing; (2) identifying AL #1 and #2, as well as secondary points along the radius bony margin (for demarcation); (3) identifying AL #3, #4, #5 and #6, as well as secondary points along the carpal bony margin (for demarcation), (4) generating the respective curves that best fit the points found in step 2 and step 4; (5) calculating distance and area.

In step 1, image enhancement was performed using two operations, namely magnification and smoothing; the latter removes pixilation as a result of enlargement. The recommended filter size of 3x3 pixels was adhered to [22]. In step 2, prior to the identification of ALs #1 and #2 on the radius bony margin (Figure 1), the WRISTJS generated a straight line through AL #1 and #2; the line was used as a guide for identifying AL #1 and #2. Secondary points were then identified along the radius bony margin (between AL #1 and #2); thereafter a curve (a 3rd order polynomial equation) was generated to best fit these points to demarcate the radius bony margin. Similar approach was applied to the identification of AL #3 to AL #6 and secondary points and demarcating the bony margins. After the ALs and demarcation lines were established, the areas, A_{MRS} and A_{LRL}, and distances, d1 to d8, were determined.

2.4 Calibration using the grid phantom

Calibration was carried out on the WRISTJS before it could be employed for distance and area measurement. The calibration method involves: (1) acquiring a DXA image of a square grid phantom; and (2) establishing a relationship between distance and area measurements (in pixels) based on the image by WRISTJS to the respective nominal values measured directly on the grid phantom. The grid was made from copper; each square has a length of 10 mm (diameters ≈0.5 mm). The grids were embedded in a block of acrylic material (240mm x 150mm x 15mm). Scanning was executed with the block placed onto the patient's couch, with the longer side parallel to the line of motion of the X-ray source. To account for any distortion in the hand image introduced by the DXA scanner, several sets of measurements of *ds* and *As*, over different parts (and sizes) of the grid, were determined with mean values calculated.

For distance calibration, a total of ten lines were drawn on the grid image (five along the width and five along the length of the block). Width-wise, each line spanned across twelve grids; length-wise, each line spanned across twenty grids. For area calibration a total of fifteen square regions were drawn on the grid image. Each square region (which covered twenty-five square grids) may overlap with another given region.

2.5 Accuracy and precision assessments using the Leeds phantom

The accuracy test compared the *ds* and *As* from a DXA image of a simple hand and wrist phantom, determined using WRISTJS (following the proposed measurement protocol described in section 2.3) with the results obtained using a direct manual method. The hand and wrist phantom (Leeds Test Objects, Leeds, United Kingdom) was constructed from acrylic and aluminum materials to mimic soft tissues and bones of the hand and wrist respectively.

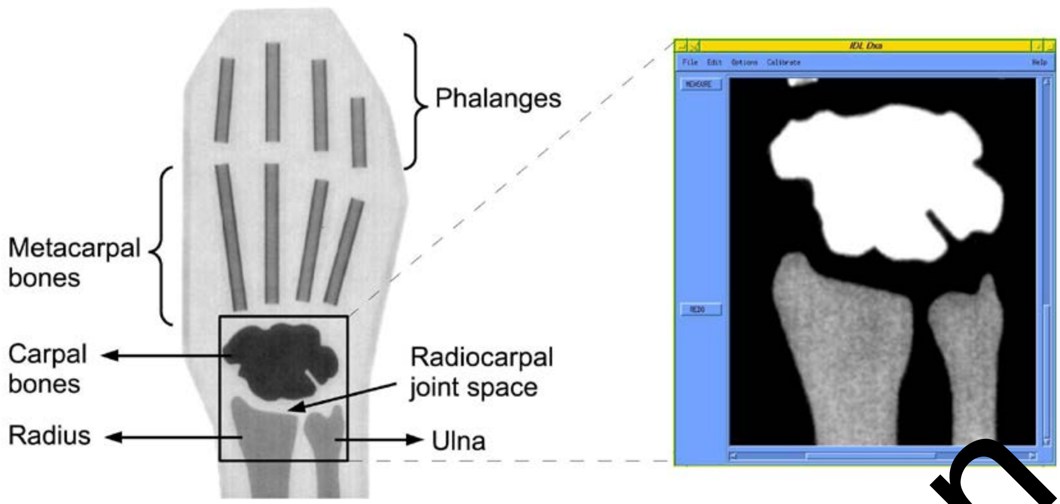

Figure 2. The hand and wrist phantom ("right hand"). There are eight aluminum rods (to model the phalanx and metacarpal of the hand) and three irregular aluminum plates (to model carpals, radius and ulna of the wrist) embedded in the acrylic material. The thickness of the main body of the phantom varies at different locations. The thickest section (≈3 cm) begins at the wrist and the thickness tapers towards the proximal end of the hand (1.5 cm). All rods and plates are less than 0.5 cm thick.

The direct manual method measured the phantom wrist JS from carefully traced drawings. A collimated light from an over-head projector (simulating the DXA imaging system) used to illuminate the phantom resulted in an outline of the bony image on a square-gridded (1mm by 1mm) tracing paper. The projected image was traced directly onto the paper using a pencil. Measurements of distances, made using a ruler (precision of ±0.01 mm), and areas (counting the number of whole and half squares) were obtained directly from the outlined image; these measured values were regarded as the expected results. In this study, measurements were only derived for d1to d4 and A_{MRS} because of difficulties in identifying the equivalent region enclosed by the lunate-radius bone on the phantom.

The accuracy of the measurement protocol was parameterised by the systematic error (SER), defined as SER = 100(*a*-*b*)/*b*, where *b* denotes the measurement quantity (distance or area) determined by WRISTJS and *a*, the corresponding measurement quantity determined by direct manual method. The precision of the measurement protocol was parameterised by the Coefficient of Variation (CV), defined as CV = 100(SD/μ), where SD is the standard deviation of a variable (i.e. distance or area) and μ_X is the mean value of the variable (i.e. distance or area).

2.6 Repeatability and reproducibility assessment of healthy and RA wrists

Two archetypal radiographs were obtained for the study: one from the right hand of a healthy subject; and the other from the right hand of a RA subject. Both subjects were from a pool of volunteers for a cross-sectional study at Woolmanhill Hospital (Aberdeen) on arthritic diseases. It was intentionally decided to evaluate only these two archetypes using the digital image analysis in order to assess the repeatability (intra-observer) and reproducibility (inter-observer) of the protocol. The DXA images were obtained at the Osteoporosis Research Unit, at the Woolmanhill Hospital. The healthy subject was a 25-year old English male volunteer. The RA patient was a 60-year old Scottish male exhibiting an advanced stage of the disease. Ethical approval was granted by the Grampian research ethics committee.

Radiographic images of the wrist were acquired using the LUNAR EXPERT-XL DXA scanner, following an imaging protocol for RA assessment. The phantom was employed to determine the accuracy of the measurement protocol. Using WRISTJS, the images were digitally magnified four times so that the radiocarpal JS was sufficiently visible to the observer. Following the proposed measurement protocol both distance and area measurement were performed.

Two experienced observers and one inexperienced observer participated in the assessment. The inexperienced observer was given instructions on how to implement the measurement protocol and was requested to make five practice attempts prior to commencing the assessment. After these practice attempts the results were immediately revealed to the participant. The aim was to provide an indication to the observer about the particular style that the operator has adopted, which could serve to reduce the range of variability.

With regards to the test for repeatability, each observer was instructed to determine two sets (repeated for fifteen times per set) of *d* and *A* measurements on two separate occasions. The interval between the first and second set was about 24 h. The reproducibility test involved only the results from the two experienced observers from the first set.

2.7 Statistical analysis

Student's t-test was employed to assess the significance of the repeatability between measurements obtained on two separate occasions for each observer. The null hypothesis (no significant difference between the mean readings of the two sets of measurements) was tested against the alternative hypothesis. For the reproducibility test, the t-test was employed to test for differences in the measurements made by two observers. The null hypothesis (no significant difference between the results obtained by the two observers) was once again tested against the alternative hypothesis. In this study, significance was defined as $P < 0.05$. The t-tests were performed using Minitab commercial software (version 16). The results of the respective *ds* and *As* were reported as means, SDs, SERs and CVs.

3 Results & Discussion

3.1 DXA images

Figure 3 shows DXA images of a healthy (right) wrist and a RA (right) wrist. The vital bones, namely: radius; ulna; scaphoid; and lunate are clearly labelled. For the healthy wrist, the fingers could be easily extended, thus resulting in the well aligned phalanges. More importantly the joint spaces of the hand, particularly those at the wrist between: the scaphoid and lunate; scaphoid and radius; and lunate and radius, are clearly visible exhibiting a high contrast image. Conversely, the RA patient found it difficult to fully extend the fingers and wrist during the imaging process, resulting in a distorted arrangement. The RA image revealed that the index finger was a somewhat bent. More importantly, many of the joint spaces that are visible in the healthy subject are not clearly visible in the RA hand. For instance, there is no clear separation distance between the lunate and scaphoid; there is a faint line between the scaphoid and radius as well as between the lunate and radius. Note that this could present difficulties when identifying the ALs and the secondary points for the measurement protocol. For this purpose, the experienced and inexperienced observers were instructed to inspect the wrist region carefully prior to executing the measurement protocol.

3.2 Accuracy and precision

Following the procedure for calibrating the WRISTJS (section 2.4), the calibration constants for converting image-related pixel numbers to units of length (mm) and area (mm^2) were found to be 1.973 pixels/mm and 3.997 pixels/mm^2 respectively (Note, these are mean values with corresponding SDs of 0.001 pixels/mm and 0.015 pixels/ mm^2).

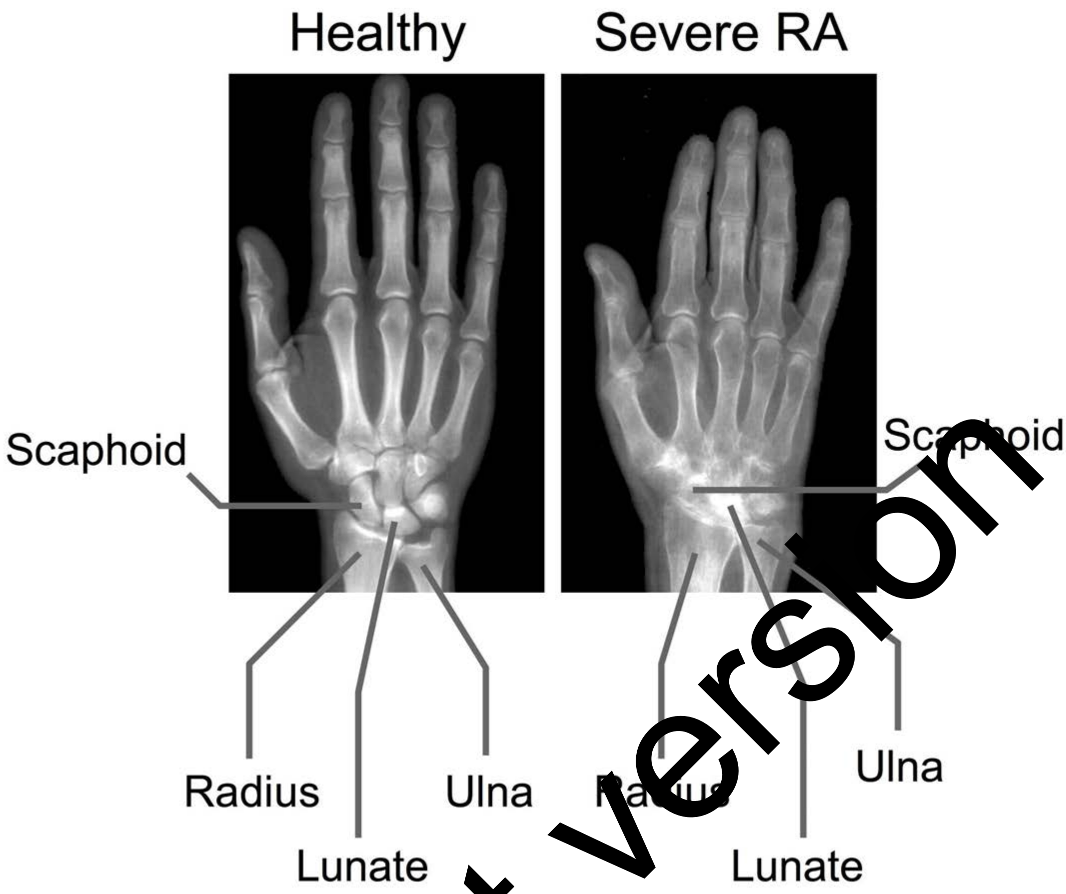

Figure 3. DXA images of a healthy wrist (left) and a RA wrist (right). The RA wrist shown here has been diagnosed as severe RA.

Table 1 lists the results of the accuracy and precision tests using the phantom. With regards to accuracy, the SER ranges from 2.45% (d1) to 7.16% (d2) for distance measurement; this returns a mean of 5.20%. The SER equals 3.49% for area measurement. As the SER is related to the difference between the values obtained using the direct method and the WRISTJS method, this study shows that the distance and area measurements obtained using WRISTJS consistently underestimated the measurements obtained using the direct method.

Table 1. Accuracy and precision tests for distance and area measurements using the phantom.

	WRISTJS	**Direct Method**$^{\nabla}$	**SER %**$^{@}$
d1	5.16 ± 0.09 mm (2.85 %)	5.42 ± 0.09 mm (1.66%)	4.83
d2	4.89 ± 0.08 mm (1.64 %)	5.27 ± 0.10 mm (1.90 %)	7.16
d3	4.60 ± 0.07 mm (1.52 %)	4.91 ± 0.16 mm (3.26 %)	6.42
d4	4.40 ± 0.08 mm (1.82 %)	4.51 ± 0.17 mm (3.77 %)	2.45
A_{MRS}	78.50 ± 1.64 mm^2 (2.09 %)	81.33 ± 5.81 mm^2 (7.14 %)	3.4[illegible]

[#] Each mean value shown in the table was determined from six repeated measurements. [@] SER % represents systematic error. Values entered under 'WRISTJS' and 'Direct method' are mean ± SD (CV %).

With regards to precision, the CV ranges from 1.52 % (d3) to 2.85 % (d1) for distance measurements obtained by the WRISTJS method and 1.66 % (d1) to 3.77 % (d4) for the direct method. Note that these return a mean of 1.96 % (WRISTJS) and 2.65 % (direct method). The area measurement reveals that the CV equals 2.09 % (WRISTJS) and 7.14 % (direct method). Overall, the results from the precision study are encouraging in that all of the precision tests show a variation (denoted by CV) of less than 10% for both area and distance. After closer scrutiny one notices that the measurement protocol implemented through WRISTJS exhibits lower variation when compared to the direct method. It can however be assumed that the precision of the direct method would improve, provided a dedicated microscope is used for both distance and area measurements. This has been exemplified by James et al [4], where such a configuration was key to securing accurate spatial measurements. Additionally, tracing paper with much smaller square grids (e.g. smaller than 1 mm^2) may be considered with the use of such a dual-configuration microscope.

Overall, these figures for the accuracy test are higher than those quoted in the literature [4, 9]. It should be emphasized that the current work performed measurements on a highly irregular “radiocarpal joint space” of the phantom. In contrast, phantoms with bone-mimicking geometries for measuring accuracy [4] should be viewed with discretion since shapes of real joint spaces are far from regular.

For the software developed by James et al. [4], it was reported that the method resulted in SERs ranging from 0.04% to 2.07% (software method versus direct method). When compared to the results in this present study (Table 1), the higher reported accuracy should not come as any surprise since the “phantom annular joint space” was measured instead. Although this led to an easier task in the measurement procedure, it was inadequate for the joint space to be represented by any form of regular geometrical shape since real joint spaces are highly irregular. For the direct method employed by James et al [4], a microscope slide calibration graticule was used, a device unavailable to the authors of this present study at the time of experimentation.

Dacre *et al.* [9] have reported that their image analysis software yielded an accuracy of less than 1%; and less than 1.3% for distance and area measurements respectively. With regards to precision, the CVs of the distance and area measurements obtained were less than 5.5%. In relation to the accuracy study, Dacre et al. [9] also employed phantoms with recognizable geometries which made the task of finding the expected values much easier.

3.3 Inter-observer effects on repeatability

The results of the study on the effects of inter-observer effects on repeatability are tabulated in Table 2. In the repeatability test for measurements taken from the healthy wrist, the CV for experienced observer 1 ranges from 6.84 % (d8) to 17.64 % (d1) for distance (mean CV = 10.87 %). For experienced observer 2, the CV ranges from 4.62 % (d8) to 12.02 % (d1) (mean CV = 7.68 %). As for the inexperienced observer, the CV ranges 6.84 % (d8) to 14.99 % (d5) (mean CV = 11.38 %). Altogether it may be concluded that the measurement of d1 resulted in the highest variability. A comparison of the average CVs for the respective observers shows that the average CV from experienced observer 2 is the lowest and that of the inexperienced observer is the highest. Overall there is no discernable difference between the results obtained by the three observers. With regards to area measurement, the CVs for experienced observer 1 are 13.84 % (A_{MRS}) and 6.4[illegible] % (A_{LRL}). For experienced observer 2, the CVs are 8.19 % (A_{MRS}) and 5.97 % (A_{LRL}). As for the inexperienced observer, the CVs are 10.76 % (A_{MRS}) and [illegible].3[illegible] % (A_{LRL}). Altogether it may be concluded that the measurement of A_{MRS} yields consistently higher variability compared to A_{LRL}. Between experienced observer 2 and the inexperienced observer, the variation in the area measurement is higher for the latter than the former. However, between experienced observer 1 and the inexperienced observer, the comparison yields mixed results: the variability associated with A_{MRS} is higher for the experienced observer 1 than the inexperienced observer but the contrary is observed for the variability associated with A_{LRL}, i.e. lower for the experienced observer 1 than the inexperienced observer.

Table 2. Inter-observer effects on the repeatability of distance and area measurement on healthy and RA wrists.

	Healthy wrist		RA wrist	
	Mean ± SD (mm)	CV (%)	Mean ± SD (mm)	CV (%)
d1* EO 1	2.05 ± 0.36	17.64	2.00 ± 0.09	4.49
EO2	3.26 ± 0.39	12.02	2.56 ± 0.17	6.54
IO	3.24 ± 0.42	12.87	2.05 ± 0.22	10.52
d2* EO1	2.29 ± 0.33	14.41	2.58 ± 0.15	5.62
EO 2	3.20 ± 0.20	6.21	2.96 ± 0.17	5.65
IO	2.79 ± 0.26	9.16	2.54 ± 0.18	7.19
d3* EO 1	2.43 ± 0.32	13.21	2.99 ± 0.20	6.61
EO 2	3.10 ± 0.21	6.61	3.22 ± 0.17	5.21
IO	2.40 ± 0.24	10.13	2.92 ± 0.22	7.[illegible]
d4* EO 1	2.47 ± 0.31	12.55	1.66 ± 0.15	9.03
EO 2	2.97 ± 0.30	10.10	2.08 ± 0.18	8.44
IO	2.09 ± 0.31	14.83	1.75 ± 0.14	8.14
d5* EO 1	2.30 ± 0.18	7.62	1.83 ± 0.16	8.60
EO 2	2.99 ± 0.25	8.34	2.09 ± 0.15	7.41
IO	2.57 ± 0.39	14.99	2.17 ± 0.30	13.68
d6* EO 1	2.47 ± 0.19	7.[illegible]4	1.67 ± 0.15	9.07
EO 2	3.22 ± 0.24	7.4[illegible]	1.99 ± 0.15	7.44
IO	2.85 ± 0.36	12.[illegible]0	1.98 ± 0.26	13.16
d7* EO 1	2.72 ± 0.1[illegible]	7.15	1.60 ± 0.15	9.13
EO 2	3.52 ± 0.2[illegible]	6.12	1.99 ± 0.15	7.51
IO	3.[illegible]7 ± [illegible].31	9.70	1.83 ± 0.20	10.86
d8* EO 1	3.07 ± 0.21	6.84	1.66 ± 0.15	9.03
EO 2	3.[illegible]9 ± 0.18	4.62	2.08 ± 0.18	8.44
IO	3.51 ± 0.24	6.84	1.75 ± 0.14	8.14
	Mean ± SD (mm^2)	CV (%)	Mean ± SD (mm^2)	CV (%)
A_{MRS}				
EO1	61.01 ± 8.44	13.84	59.95 ± 3.45	5.75
EO 2	91.57 ± 7.50	8.19	68.08 ± 4.63	6.80
IO	71.28 ± 7.67	10.76	59.70 ± 5.95	9.97
A_{LRL}				
EO1	45.62 ± 2.96	6.49	23.02 ± 1.81	7.85
EO 2	56.14 ± 3.35	5.97	26.50 ± 1.64	6.19

IO	50.36 ± 4.70	9.34	25.40 ± 2.93	11.54

* d1 to d8 refers to the locations where the distance measurements were made. A_{MRS} and A_{LRL} denote the area of the radioscaphoid and radiolunate regions, respectively. EO and IO denote experienced and inexperienced observers, respectively. Each observer carried out fifteen repeated measurements at each location.

With regards to measurements taken from the RA wrist, the CV for experienced observer 1 ranges from 4.49 % (d1) to 9.07 % (d5; the next closest value is 9.03 % for d8) for distance (mean CV = 7.70 %). For experienced observer 2, the CV ranges from 5.21 % (d3) to 8.44 % (d4) (mean CV = 7.08 %). As for the inexperienced observer, the CV ranges from 7.19 % (d2) to 13.68 % (d[illegible]) (mean CV = 9.92 %). Unlike the results for the healthy wrist where d1 appears to contribute to high variability, regardless of observer, here it appears that the degree of variability is dependent on the observer. A comparison of the average CVs for the respective observers shows that there is no appreciable difference in the variability between experienced observer 1 and 2; the inexperienced observer reveals the highest variability. With regards to area measurement, the CVs for experienced observer 1 are 5.75 % (A_{MRS}) and 7.85 % (A_{LRL}). For experienced observer 2, the CVs are 6.80 % (A_{MRS}) and 6.1[illegible] % (A_{LRL}). As for the inexperienced observer, the CVs are 9.97 % (A_{MRS}) and 11.54 % (A_{LRL}). To some extent, similar to the healthy wrist, it may be concluded that the measurement of A_{MRS} yields consistently higher variability compared to A_{LRL}. This applies to experienced observer 1 and the inexperienced observer; for experienced observer 2, the variability in the measurement of A_{MRS} is comparable to A_{LRL}. Nevertheless, between the experienced observers and the inexperienced observer, the variation in area measurement is higher for the latter than the former.

It is observed that the variation is generally higher for distance and for area taken from the healthy subject when compared to the RA subject. This may be attributed to the difficulty in identifying the AL #5 on the healthy image. Initial inspection of the image of the healthy wrist revealed that there was no perceivable radius leading edge "intersecting" the medial scaphoid edge. All observers had been instructed on where to measure AL #5, which adhered to the original definition, yet did not contrast with the surrounding features. As a result, it is possible that the inconspicuous appearance of this point contributed to the difficulty in executing repeated measurements. No such difficulty was registered by the observers for the RA image.

Overall, it appears that the repeatability is highest for experienced observer 2; it is debatable whether the measurements taken by experienced observer 1 are any more repeatable than the measurements taken by the inexperienced observer. One may envisage that repeatability could improve if measurements are performed more judiciously (as observed for experienced observer 2). It was noted that experienced observer 2 dedicated approx. 1.5 hours to take all fifteen measurements. However, the time dedicated by experienced observer 1; and by the inexperienced observer was much shorter (approx. 40 minutes for fifteen repeated measurements).

3.4 Intra-observer effects on repeatability

The results for the intra-observer effects on the repeatability of distance and area measurements made on a RA wrist are listed in Table 3. Note that the first and second sets of results were obtained in two separate imaging sessions (24 hours apart).

Table 3. Intra-observer effects on the repeatability of the distance and area measurement for the RA wrist.

		EO 1	EO 2	IO
d1*	1st set (mm)	2.00 ± 0.09	2.56 ± 0.17	2.05 ± 0.22
	2nd set (mm)	2.08 ± 0.13	2.54 ± 0.13	2.20 ± 0.14
	p	0.069	0.70	0.03
	T	-1.90	0.38	-2.41
d2*	1st set (mm)	2.58 ± 0.15	2.96 ± 0.17	2.54 ± .18
	2nd set (mm)	2.72 ± 0.13	2.95 ± 0.14	2.69 ± .18
	p	0.0072	0.92	0.026
	T	-2.91	0.11 (NS)	-2.48
d3*	1st set (mm)	2.99 ± 0.20	3.22 ± 0.17	[illegible].92 ± 0.22
	2nd set (mm)	3.16 ± 0.16	3.22 ± 0.16	3.04 ± 0.2[illegible]
	p	0.013	1.00	[illegible].068
	T	-2.68	-0.01 (NS)	[illegible]1.98
d4*	1st set (mm)	3.22 ± 0.23	3.36 ± 0.17	3.15 ± 0.26
	2nd set (mm)	3.38 ± 0.17	3.3[illegible] ± 0.16	3.26 ± 0.33
	p	0.031	0.99	0.162
	T	-2.28	-0.01 (NS)	-1.48
d5*	1st set (mm)	1.83 ± 0.16	2.09 ± 0.16	2.17 ± 0.30
	2nd set (mm)	1.87 ± 0.16	1.95 ± 0.20	2.00 ± 0.27
	p	0.54	0.049	0.12
	T	-0.62	2.06	1.59
d6*	1st set (mm)	1.6[illegible] ± 0.15	1.99 ± 0.15	1.98 ± 0.26
	2nd set (mm)	1.70 ± 0.15	1.87 ± 0.16	1.81 ± 0.23
	p	0.60	0.039	0.065
	T	-0.54 (NS)	2.17	1.92
d7*	1st set (mm)	1.60 ± 0.15	1.99 ± 0.15	1.83 ± 0.20
	2nd set (mm)	1.63 ± 0.15	1.89 ± 0.12	1.68 ± 0.17
	p	0.61	0.057	0.035
	T	-0.51 (NS)	1.99	2.21
d8*	1st set (mm)	1.66 ± 0.15	2.08 ± 0.18	1.75 ± 0.14
	2nd set (mm)	1.69 ± 0.14	2.02 ± 0.10	1.68 ± 0.17
	p	0.59	0.31	0.25
	T	-0.55 (NS)	1.03	1.18
A_{MRS}	1st set (mm^2)	59.95 ± 3.45	68.08 ± 4.63	59.70 ± 5.95
	2nd set (mm^2)	63.13 ± 2.54	67.46 ± 4.42	58.74 ± 4.08
	p	0.027	0.744	0.549

	T	-2.47	0.33 (NS)	0.61
A_{LRL}	1st set (mm^2)	23.02 ± 1.81	26.50 ± 1.64	25.40 ± 2.93
	2nd set (mm^2)	23.33 ± 1.87	25.50 ± 1.72	24.00 ± 1.99
	p	-0.698	0.105	0.138
	T	-0.40 (NS)	1.74	1.58

* d1 to d8 refers to the locations where the distance measurement were made. T refers to the t-test statistic. NS denotes not statistically significant ($p > 0.05$). EO and IO denote experienced and inexperienced observers, respectively. The 1st and 2nd sets of values were taken in two separate sessions.

Overall, there were no significant differences ($p > 0.05$) in the distance measurement between the 2 sets of results in 5 out of 8 locations (i.e. d1 to d8). Experienced observer 2 yielded marginally better results, yet there were no significant differences ($p > 0.05$) in the distance measurement in 7 out of 8 locations. The inexperienced observer obtained results which showed no significant difference ($p > 0.05$) in 5 out of 8 locations (similar to experienced observer 1). Given the mixed results, there is not a strong basis to conclude that intra-observation has an effect on distance measurement. From a location perspective (i.e. d1 to d8), these observations suggest that there is no systematic trend for accepting the null hypothesis and the results were independent of location. In other words, location has no effect on the results.

With regards to the area measurement, the results of experience observer 1 show that there was no significant difference ($p > 0.05$) in the measurement of A_{LRL}, but not for A_{MRS}. However, the results of experience observer 2 and the inexperienced observer all yielded no significant difference ($p > 0.05$) in the measurement of A_{MRS} and A_{LRL}. With the exception of the experienced observer 1 (where A_{MRS} is concerned), it may be concluded that intra-observer effects have little effect on area measurement.

3.5 Inter-observer effects on reproducibility

The results for the inter-observer effects on the reproducibility of distance and area measurements made on a RA wrist are listed in Table 4, by the two experienced observers. With regards to distance measurement, there were no significant differences ($p > 0.05$) in the measurement between the results of the two observers in 4 out of 8 locations. No significant differences in measurement were found at locations d3 to d6. With regards to area measurement, there was no significant difference ($p > 0.05$) in the measurement of A_{MRS} by the two experienced observers, but this is not true for A_{LRL}. Additionally, the conclusions established by the t-tests are consistent with the results shown for the respective CVs; the CVs are large where significant differences are observed (but small where no significant differences are observed).

What could be a possible cause of the discrepancies in the results from both observers? Here we noted that one of the observers autonomously applied additional image enhancement methods (e.g. color and contrast) before taking every measurement. In hindsight, one might then suggest that all observers should have adopted the same image enhancement methods. Additionally, for distance measurement, the CVs at d1; d5; d6; d7; and d8 are all ≈10 %. This may be because the curves fitted to the bony margin of the radioscaphoid and radiolunate differ at the end points to the curves produced by the observers for ALs #1, #2, #5 and #6.

Table 4. Inter-observer effects on the reproducibility of the distance and area measurement for the RA wrist.

	Mean ± SD (mm) EO 1	**Mean ± SD (mm) EO 2**	**Mean ± SD (mm) (EO 1 & 2)**	**Overall CV% (EO 1 & 2)**	**p**	**T**
d1*	2.08 ± 0.13	2.54 ± 0.13	2.31 ± 0.27	11.63	0	-11.21
d2*	2.72 ± 0.13	2.95± 0.14	2.84 ± 0.16	6.20	0	-4.58
d3*	3.16 ± 0.16	3.22 ± 0.16	3.19 ± 0.16	4.89	0.305	-1.06
d4*	3.38 ± 0.17	3.36 ± 0.16	3.37 ± 0.17	4.94	0.67	0.44
d5*	1.87 ± 0.16	1.95 ± 0.20	1.91 ± 0.18	9.32	0.192	-1.37
d6*	1.70 ± 0.15	1.87 ± 0.16	1.79 ± 0.18	9.85	0.06	-3.26
d7*	1.63 ± 0.15	1.89 ± 0.12	1.76 ± 0.19	10.53	0	-6.76
d8*	1.69 ± 0.14	2.02 ± 0.10	1.85 ± 0.21	11.22	0	-14.04
	Mean ± SD (mm^2) EO 1	Mean ± SD (mm^2) EO 2	Mean ± SD (mm^2) (EO 1 & 2)	Overall CV% (EO 1 & 2)	p	T
A_{MRS}	63.13 ± 2.56	67.46 ± 4.42	65.30 ± 4.18	6.4	0.07	-3.13
A_{LRL}	23.33 ± 1.87	25.50 ± 1.72	24.42 ± 2.09	8.53	0.02	-3.92

* d1 to d8 refers to the locations where the distance measurements were made (see Chapter 3). EO and IO denote experienced and inexperienced observers, respectively. T is the t-test statistic. Measurements at all locations were repeated for fifteen times for every observer.

3.6 Suggestion for further studies

In principle, all steps in the protocol must be strictly adhered so as to ensure the highest level of consistency in the results. In practice, discrepancies do arise from inter and intra-observer effects. Given that only three observers were used in this preliminary study, it is not clear if one could conclude that experienced observers fared better than inexperienced observer. Nevertheless, it must be pointed out that the CVs from the experienced observers are ≈10% or less (with one or two exceptions). In comparison, the variabilities found in the repeatability study of James et al. [4] were generally twice as much as those reported in this present study, although James and co-workers had reportedly employed a much more sophisticated thresholding method in their computerized image analysis method.

At present, visual inspection is employed to decide the optimal order for the fitting of the polynomial function to the points demarcating the bony margins. It may be necessary to quantify the optimization method in order to provide a more accurate determination of the order of fit. A method suggested by Dacre *et al.* [5] to determine the optimal number of points to fit a curve may be modified for the determination of the order of fit. Alternatively, a method involving both curve-fitting and thresholding of edges may be employed as suggested by James et al. [4]. This method may ensure a more accurate delineation of the bony margin.

Currently, the number of points implemented to identify the bony outline of the radiocarpal JS are chosen out of convenience and are also limited by the size of the joint space. In turn, the number of points used also determines the order of the polynomial function. In consideration of these two variables (number of points and order of the polynomial function), it would be interesting to determine quantitatively the optimal number of points (for any fixed order) needed to achieve consistent distance and area measurement.

The reliability of the anatomical locations has not been properly addressed in this work. Preliminary results from the study of inter-observer effects on reproducibility (section 3.5) implicate the contribution of the positions of anatomical locations at #1, #2, #5 and #6 to the discrepancy in the readings. It may warrant further investigation to find out how the position of each of these locations will influence the final results.

The method to select locations within each compartment of the radiocarpal JS for distance measurement may require further modifications. A limitation of the method is that it relies on the image's Cartesian coordinate axes. Here, the vertical axis of the image (as seen on a display screen, with the hand's axis upright) is in the direction of motion of the X-ray source. The locations for distance measurements are then found with respect to the horizontal axis (section 2.2). The standard scanning protocol requires that the hand be positioned with its arm's axis parallel to the direction of the scan motion for consistent results [5]. In practice it cannot always be ensured that the arm's axis is precisely positioned is such a manner. In this respect, results obtained on subsequent scans may not be comparable or reproducible. A viable alternative method would be to determine equally spaced points along the fitted radius curve using one end of the curve as the origin.

The accuracy tests conducted in this work may warrant further improvement. A dedicated microscope (as already suggested by James et al. [4]) and a tracing paper designed with much smaller square grids are necessary prerequisites for establishing a more precise (smaller variability) value of the expected value obtained from the direct manual method.

In the repeatability test to compare readings obtained on two separate occasions (section 3.4), preliminary results appeared to suggest that additional pre-processing image enhancement methods led to a smaller variability. Further tests involving more experienced observers employing similar additional pre-processing image enhancement methods will be needed to assess to this claim.

In the reproducibility test to investigate the precision of WRISTJS, preliminary results suggest that different pre-processing image enhancement methods between two experienced observers may have led to different final results (section 3.5). Future reproducibility investigations would have observers employing similar pre-processing image enhancement methods to ensure that the results are more comparable.

As in all protocols related to radiographic assessment, there should be checks along the way to minimise inconsistency. For instance, a longer duration could be designated (i.e. one week, see Harvey *et al.* [13]) between two assessment sessions. Also, we should ask if the results could be improved. For instance, reduce the variability by providing feedback to the operator immediately after each session so as to help the operator reflects on their approach.

Anatomical locations #5 and #6 are found along the medial scaphoid and lateral lunate respectively. These are determined by the "intersection" of the radius leading edge with the respective carpal bone margin. "Intersection" is a misnomer here because the radiographic image is a projection image and basically shows the overlapping of the radius bone with the carpal bones. Consequently, it must be emphasized that in severe RA wrists the radial bone may overlap extensively with the carpal bones. In order for complete delineation of the carpal bony margin there is a thirteenth point which lies mid-way between the carpal tangential locations #3 and #4. A point within this joint space is necessary to ensure that the curve generated will not overlap with the radius curve in the event that the joint space becomes too narrow.

The accuracy of the measurement is reflected by the systematic errors in the measurement (section 3.2). WRISTJS includes a calibration procedure which attempts to correct for systematic errors arising from image distortion (caused by the geometry of the fan-beam X-ray). This is done by taking mean calibration measurements obtained from different parts of an image of a grid phantom. With regards to the method of measurement, the polynomial curve-fitting technique for delineating bony margins should be re-visited. Determining the optimal order of the polynomial function by visual inspection is key to minimizing systematic errors arising from the curve-fitting technique [5, 9].

4 Conclusion

We have demonstrated the feasibility of a digital image processing protocol for measuring JS in the radiocarpal region of the wrist from radiographic images. The protocol was implemented by a computer algorithm, featuring an interactive user interface, executed on the IDL platform. The overall protocol comprises the image processing stage; the identification of anatomical locations; as well as secondary points, within the JS and the determination distance and area. A phantom was used to assess the accuracy and precision of the protocol. Preliminary studies were carried out on DXA images of a healthy wrist and a RA wrist. Three observers participated in this study. Tests were carried out to study the effects of inter-observer on the repeatability of the protocol, as well as the effects of intra-observer on both the repeatability and reproducibility of the protocol. For the accuracy and precision study; and for the repeatability and reproducibility study the variabilities were found to be about 10% or less. These findings, in addition to the versatility and simplicity of the digital image analysis protocol, lend to the potential for further studies such as using the protocol to complement the acquisition of bone mineral density data derived from DXA for diagnosing the progression of the RA in patients.

5 Acknowledgements

The authors thank Jo Norris, David Reid, and David Hukins for supplying the DXA images; and also for assisting with use of the DXA facility at Woolmanhill Hospital, Aberdeen.

Conflicts of interest

The authors declare no conflict of interest.

References

[1] K. Trieb, "Treatment of the wrist in rheumatoid arthritis," *The Journal of hand surgery,* vol. 33, pp. 113-123, 2008.

[2] D. P. M. Symmons, E. M. Barrett, C. R. Bankhead, D. G. L. Scott, and A. J. Silman, "The incidence of rheumatoid arthritis in the United Kingdom: results from the Norfolk Arthritis Register," *Rheumatology,* vol. 33, pp. 735-739, 1994.

[3] J. H. Kellgren and J. S. Lawrence, "Radiological assessment of osteo-arthrosis," *Annals of the rheumatic diseases,* vol. 16, p. 494, 1957.

[4] M. F. James, G. Heald, J. H. Shorter, and R. A. Turner, "Joint space measurement in hand radiographs using computerized image analysis," *Arthritis & Rheumatism,* vol. 38, pp. 891-901, 1995.

[5] J. T. Sharp, J. C. Gardner, and E. M. Bennett, "Computer-based methods for measuring joint space and estimating erosion volume in the finger and wrist joints of patients with rheumatoid arthritis," *Arthritis & Rheumatism,* vol. 43, pp. 1378-1386, 2000.

[6] S. Boini and F. Guillemin, "Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages," *Annals of the rheumatic diseases,* vol. 60, pp. 817-827, 2001.

[7] P. A. Ory, "Interpreting radiographic data in rheumatoid arthritis," *Annals of the rheumatic diseases,* vol. 62, pp. 597-604, 2003.

[8] J. C. Buckland-Wright, "Quantitative radiography of osteoarthritis," *Annals of the rheumatic diseases*, vol. 53, p. 268, 1994.

[9] J. E. Dacre, J. S. Coppock, K. E. Herbert, D. Perrett, and E. C. Huskisson, "Development of a new radiographic scoring system using digital image analysis," *Annals of the rheumatic diseases*, vol. 48, pp. 194-200, 1989.

[10] J. C. Buckland-Wright, I. Carmichael, and S. R. Walker, "Quantitative microfocal radiography accurately detects joint changes in rheumatoid arthritis," *Annals of the rheumatic diseases*, vol. 45, pp. 379-383, 1986.

[11] J. C. Buckland-Wright, D. G. Macfarlane, S. A. Williams, and R. J. Ward, "Accuracy and precision of joint space width measurements in standard and macroradiographs of osteoarthritic knees," *Annals of the rheumatic diseases*, vol. 54, pp. 872-880, 1995.

[12] J. C. Buckland-Wright, "A new high-definition microfocal X-ray unit," *The British journal of radiology*, vol. 62, pp. 201-208, 1989.

[13] S. B. Harvey, K. M. Hutchison, E. C. Rennie, D. W. Hukins, and D. M. Reid, "Comparison of the precision of two vertebral morphometry programs for the lunar EXPERT-XL imaging densitometer," *The British journal of radiology*, vol. 71, pp. 388-398, 1998.

[14] K. L. Goh, S. C. Liew, and B. H. Hasegawa, "Energy dependent systematic errors in dual-energy X-ray CT," in *Nuclear Science Symposium and Medical Imaging Conference Record, 1995., 1995 IEEE*, 1995, pp. 1587-1591.

[15] K. L. Goh, S. C. Liew, and B. H. Hasegawa, "Correction of energy-dependent systematic errors in dual-energy X-ray CT using a basis material coefficients transformation method," *IEEE Transactions on Nuclear Science*, vol. 44, pp. 2419-2424, 1997.

[16] L. Yu, S. Leng, and C. H. McCollough, "Dual-energy CT–based monochromatic imaging," *American Journal of Roentgenology,* vol. 199, pp. S9-S15, 2012.

[17] M. M. Haara, J. P. A. Arokoski, H. Kröger, A. Kärkkäinen, P. Manninen, P. Knekt, *et al.*, "Association of radiological hand osteoarthritis with bone mineral mass: a population study," *Rheumatology,* vol. 44, pp. 1549-1554, 2005.

[18] M. Østergaard, B. Ejbjerg, and M. Szkudlarek, "Imaging in early rheumatoid arthritis: roles of magnetic resonance imaging, ultrasonography, conventional radiography and computed tomography," *Best Practice & Research Clinical Rheumatology,* vol. 19, pp. 91-116, 2005.

[19] M. Roberts, T. F. Cootes, and J. E. Adams, "Vertebral morphometry: semiautomatic determination of detailed shape from dual-energy X-ray absorptiometry images using active appearance models," *Investigative radiology,* vol. 41, pp. 849-859, 2006.

[20] G. Guglielmi, D. Diacinti, C. Van Kuijk, F. Aparisi, C. Krestan, J. E. Adams, *et al.*, "Vertebral morphometry: current methods and recent advances," *European radiology,* vol. 18, pp. 1484-1496, 2008.

[21] P. Peloschek, G. Langs, M. Weber, J. Sailer, M. Reisegger, H. Imhof*, et al.*, "An Automatic Model-based System for Joint Space Measurements on Hand Radiographs: Initial Experience 1," *Radiology,* vol. 245, pp. 855-862, 2007.

[22] W. K. Pratt, *Digital Image Processing: PIKS Scientific Inside*: Wiley, 2007.