Validation of automatic joint space width measurements in hand radiographs in rheumatoid arthritis

Olga Schenk
Yinghe Huo
Koen L. Vincken
Mart A. van de Laar
Ina H. H. Kuper
Kees C. H. Slump
Floris P. J. G. Lafeber
Hein J. Bernelot Moens

SPIE.

Validation of automatic joint space width measurements in hand radiographs in rheumatoid arthritis

Olga Schenk,[a,*] **Yinghe Huo,**[b,c] **Koen L. Vincken,**[b] **Mart A. van de Laar,**[d] **Ina H. H. Kuper,**[d] **Kees C. H. Slump,**[a]
Floris P. J. G. Lafeber,[c] **and Hein J. Bernelot Moens**[e]
[a]University of Twente, MIRA Institute for Biomedical Technology and Technical Medicine, Drienerlolaan 5, 7522 NB, Enschede, The Netherlands
[b]University Medical Center Utrecht, Image Sciences Institute, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands
[c]University Medical Center Utrecht, Department of Rheumatology and Clinical Immunology, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands
[d]Department of Rheumatology, Medisch Spectrum Twente, Koningsplein 1, 7512 KZ, Enschede, The Netherlands
[e]Department of Rheumatology, Ziekenhuisgroep Twente, Geerdinksweg 141, 7555 DL, Hengelo, The Netherlands

Abstract. Computerized methods promise quick, objective, and sensitive tools to quantify progression of radiological damage in rheumatoid arthritis (RA). Measurement of joint space width (JSW) in finger and wrist joints with these systems performed comparable to the Sharp–van der Heijde score (SHS). A next step toward clinical use, validation of precision and accuracy in hand joints with minimal damage, is described with a close scrutiny of sources of error. A recently developed system to measure metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints was validated in consecutive hand images of RA patients. To assess the impact of image acquisition, measurements on radiographs from a multicenter trial and from a recent prospective cohort in a single hospital were compared. Precision of the system was tested by comparing the joint space in mm in pairs of subsequent images with a short interval without progression of SHS. In case of incorrect measurements, the source of error was analyzed with a review by human experts. Accuracy was assessed by comparison with reported measurements with other systems. In the two series of radiographs, the system could automatically locate and measure 1003/1088 (92.2%) and 1143/1200 (95.3%) individual joints, respectively. In joints with a normal SHS, the average (SD) size of MCP joints was 1.7 ± 0.2 and 1.6 ± 0.3 mm in the two series of radiographs, and of PIP joints 1.0 ± 0.2 and 0.9 ± 0.2 mm. The difference in JSW between two serial radiographs with an interval of 6 to 12 months and unchanged SHS was 0.0 ± 0.1 mm, indicating very good precision. Errors occurred more often in radiographs from the multicenter cohort than in a more recent series from a single hospital. Detailed analysis of the 55/1125 (4.9%) measurements that had a discrepant paired measurement revealed that variation in the process of image acquisition (exposure in 15% and repositioning in 57%) was a more frequent source of error than incorrect delineation by the software (25%). Various steps in the validation of an automated measurement system for JSW of MCP and PIP joints are described. The use of serial radiographs from different sources, with a short interval and limited damage, is helpful to detect sources of error. Image acquisition, in particular repositioning, is a dominant source of error. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JMI.3.4.044502]

1 Introduction

Joint space narrowing (JSN) and bone erosions are important outcome measures in rheumatoid arthritis (RA). While MRI and ultrasound are suited to measure actual disease activity, conventional radiographs are widely used to assess structural damage. In clinical research, radiological outcome is often expressed with the semiquantitative Sharp–van der Heijde-score (SHS).[1] This scoring system is not well suited to clinical practice or the analysis of large datasets, because it is time consuming and subject to inter- and intrascorer variability.[1] To overcome these limitations, computerized methods were proposed for the measurement of joint space width (JSW) of finger joints.[2–7] These systems promise elimination of subjective issues and a continuous measurement scale resulting in a higher sensitivity and improved reproducibility.[4,8–14]

Automated JSW measurements depend on the quality of radiological equipment, variation in positioning and exposure of the joint, and on the measurement software.[9] Evaluation of these measurements, aiming at a precision of <0.3 mm,[4] is notoriously difficult. Several studies have reported good discriminatory ability of automated JSW measurement compared to the SHS_{JSN}.[13–16] A few studies of automated measurements found variation in JSW in the healthy population, demonstrating that decrease in JSW of metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joint with aging is normal. Against this background noise, it is important, as Finckh indicated, to evaluate the diagnostic performance of automated scoring systems between consecutive radiographs.[13]

In this study, we present a method to validate the accuracy of an automated JSW measurement system. The process includes systematic evaluation of the sources of errors, from image acquisition to automated measurements, in order to define steps for further improvement of these systems. In contrast to

*Address all correspondence to: Olga Schenk, E-mail: o.schenk@utwente.nl

earlier studies of automated JSW measurements, we used serial images without significant progression of damage. We assumed that the absence of change provides the best background to detect measurement errors related to image acquisition and/or software. In order to assess the impact of image acquisition, we used automated measurements on images from a multicenter study made with conventional radiography and standardized digital images from a single hospital.

2 Patients and Methods

Digital copies of hand radiographs of early RA patients from two studies were obtained. The randomized controlled combinatie-therapie bij reumatoïde artritis (COBRA) trial in new onset RA was performed between 1993 and 1997.[17] Conventional radiographic assessments, obtained every 6 months, had been digitized using a pixel size of 0.1×0.1 mm. The second dataset originates from the Dutch Rheumatoid Arthritis Monitoring (DREAM) remission induction cohort. This observational study in early RA was performed between 2006 and 2009. Radiographic assessments were obtained every 6 months with up-to-date digital equipment with pixels size of 0.16×0.16 mm and 0.1×0.1 mm.[18] All images were anonymized and randomly analyzed. Results of the scoring of the radiographs according to the SHS_{JSN}-method[17] were retrieved from the original study data.

Using a power analysis (two-tailed, 99% confidence interval, $\sigma = 0.3$), we estimated that at least 60 pairs of consecutive images from each dataset would comprise sufficient variation for this study. Radiographs were selected on the following criteria: availability of two consecutive (interval between 6 and 12 months) hand radiographs with known SHS_{JSN}; from the COBRA trial: pairs with all joint margins delineated correctly with digital methods in an earlier study;[4] from the DREAM cohort radiographs from a single hospital, using a constant acquisition protocol during the study. To enable adequate comparison, images with pixel size other then 0.1×0.1 mm were excluded. In line with previous studies for automated scoring, our analyses were restricted to MCP 2-5 and PIP 2-5.

3 Automated Joint Space Width Measurement

The JSW of MCP and PIP joints was measured by a single operator (YH) blinded for the SHS results. The method is described elsewhere in detail.[5,16] Briefly, as a first step, the hand outline is extracted. The next step is detection of the finger midlines followed by localization of the joints using geometric relationships of the fingers. The joint span is located using the sharpest convex corners of the head of the phalanx, and its margins are determined. The JSW is calculated as the average Euclidean distance in mm from all points on the upper margin to the lower margin and vice versa, within the middle 60% of the joint span. During the measurement process, each joint is visually checked. When the detected joint location was correct but the margin was (partly) inaccurate, the margin was adapted by the operator. Obvious misfit of joint and/or margins was marked as an algorithm malfunction by the operator. These joints were excluded from further analysis.

4 Analysis

The success of automated analysis was calculated as the percentage of joints that could be measured successfully, allowing minimal intervention by the operator. To judge the external

validity of the JSW in mm, we compared the average results with comparable measurements reported in the literature.

Precision of the measurements is based on the assumption that two consecutive radiographs with the same SHS_{JSN} should have the same JSW in mm. When there was no change in SHS assessed by human observers, a difference between the two measurements (ΔJSW) of more than 0.2 mm was regarded as incorrect. For mild progression of JSN ($\Delta SHS_{JSN} = 1$), we defined ΔJSW as correct between -0.4 and -0.2 mm; for ($\Delta SHS_{JSN} = 2$), we chose a ΔJSW ≤ -0.3 mm. The percentage of joints that showed no discrepancy according to these definitions and the standard deviation of each dataset with respect to the associated SHS_{JSN} and ΔSHS_{JSN} (precision) were calculated. To detect the causes of discrepant results, the joint margins marked by the software were presented to two rheumatologists (H. K. and H. M.), experienced in scoring radiographs. They independently classified the cause of the discrepancy as repositioning of the hand, inconsistent exposure, or incorrect delineation of the joints by the software.

5 Results

In the COBRA set, SHS scores were incomplete for 2 of 80 pairs (baseline and follow-up), leaving 78 pairs of radiographs for further analysis. A schematic of the exclusion and failures is presented in Fig. 1. In just four pairs of radiographs from the DREAM cohort, the pixel size appeared to be more than 0.1 mm and was thus excluded, leaving 76 pairs for measurements. The computerized method recognized joints and performed measurements of JSW automatically in 68 of the 78 sets of radiographs (87%) from the COBRA set. In 10 pairs, the software was unable to recognize both hands, due to the digitization (2 with smudged image), positioning of the joints (6) and failure of the software (2). In the remaining 68 consecutive paired hands, 544 MCP and 544 PIP joints (1088 individual joints) were measured. In the DREAM dataset, the automated method failed in a single MCP joint due to abnormal positioning of the hand leaving 75 (99%) successfully measured pairs of digital radiographs (600 MCP and 600 PIP joints) for analysis. There was no damage in 1188/1248 (95%) of the MCP and PIP joints in the COBRA study and in 1129/1200 (94%) in the DREAM cohort.

In 1003/1088 (92.2%) individual joints in the COBRA set, the software could automatically measure the JSW. In 51 (4.7%), it failed to define correct margins and in 26 (2.4%) to determine an accurate joint span. These could be corrected by the operator. Eight joints (1.5%) were excluded from further analysis, because recognition failure could not be corrected. Causes for these failures were a collapsed joint (1), abnormal joint margin (1), finger overlap (1), finger tissue overlap (2), and contrast issues (3). This left 536 joint pairs for further analysis. Mean and standard deviation of JSW in mm are given per SHS_{JSN} in Table 1.

In the DREAM images, the software was successful without human intervention in 1143/1200 (95.3%) individual joints. Obvious visual misfits of the margins in 27 (2.3%) and joint span in 19 (1.6%) were corrected by the operator. Eleven (0.9%) joints were not measurable with the system due to contrast issues, leaving 589 consecutive joint pairs for further analysis. Mean and standard deviation are given per SHS_{JSN} in Table 1.

To estimate the accuracy of the system under scrutiny, the average results of the measurements in mm for all MCP and PIP joints are compared to published results from other
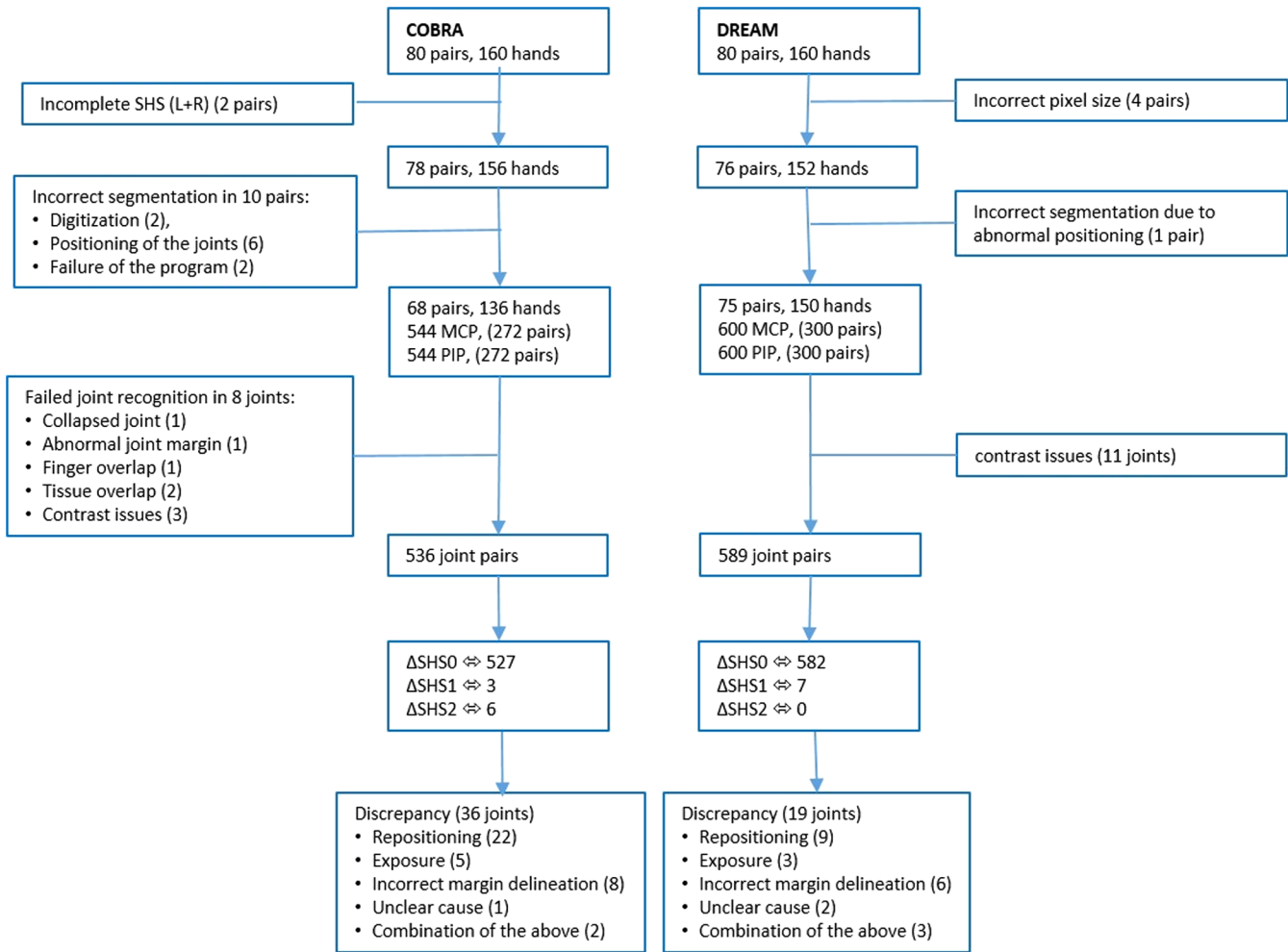
Fig. 1 Schematic of selected images, exclusions for various reasons, and success and reasons for measurement failures.

Table 1 Automated joint space width measurements in MCP and PIP joints in the two series of images related to the Sharp–van der Heijde scores (mean ± standard deviation per SHS).

| SHS_{JSN} | COBRA (1072 joints) | | DREAM (1178 joints) | |
	MCP (mm)	PIP (mm)	MCP (mm)	PIP (mm)
0	1.7 ± 0.2 ($N = 504$)	1.0 ± 0.2 ($N = 527$)	1.6 ± 0.3 ($N = 541$)	0.9 ± 0.2 ($N = 570$)
1	1.6 ± 0.2 ($N = 4$)	0.9 ± 0.2 ($N = 5$)	1.3 ± 0.2 ($N = 43$)	0.7 ± 0.2 ($N = 14$)
2	1.5 ± 0.2 ($N = 18$)	0.8 ± 0.1 ($N = 4$)	1.1 ± 0.1 ($N = 4$)	0.6 ± 0.0 ($N = 2$)
3	1.1 ± 0.2 ($N = 8$)	1.1 ± 0.1 ($N = 2$)		0.6 ± 0.1 ($N = 4$)

populations and systems in Table 2. This table illustrates that there is an overall agreement with measurements in other studies.[3,10,19–21] The JSW of MCP joints with an SHS of one or more in our study matches the measurements by van 't Klooster in osteoarthritis (OA).[3]

As a result of the short interval between consecutive images in early RA, the majority of joints had an unchanged SHS_{JSN} ($\Delta SHS_{JSN} = 0$) (Table 3). In these unchanged joints, for both COBRA and DREAM, the mean and standard deviation for MCP and PIP joints were 0.0 ± 0.1 mm. This agreement at group level conceals individual joints showing a mismatch, in some with a ΔJSW up to ± 0.5 mm. The precision of the software for unchanged MCP and PIP joints is plotted in Fig. 2 and suggests a better performance in the DREAM compared to the COBRA images. Progression of JSN was scored "1" in 11 joints ($\Delta SHS_{JSN} = 1$, COBRA: $N = 4$, DREAM remission induction: $N = 7$: Table 3). Mean automated measurements in these were 0.0 ± 0.2 mm (COBRA) and 0.0 ± 0.1 mm (DREAM). Six

**Table 2** JSW in MCP and PIP joints reported in various study populations.

Study and population	Diagnosis	MCP	PIP
COBRA with SHS = 0	Early RA	(N = 504)	(N = 527)
(N = 78)		1.7 ± 0.2 mm[a]	1.0 ± 0.2 mm
DREAM with SHS = 0	Early RA	(N = 541)	(N = 570)
(N = 76)		1.6 ± 0.3 mm[a]	0.9 ± 0.2 mm
Pfeil et al.[20]	Hand trauma	(N = 4345)	(N = 4345)
Females (N = 351)		1.6 ± 0.3 mm	1.1 ± 0.2 mm
Males (N = 518)		1.8 ± 0.4 mm	1.3 ± 0.3 mm
Goligher et al.[18]	Early RA	1.63 (1.54, 1.72) mm[b]	0.99 (0.94, 1.04) mm 1.28 (1.18, 1.38) mm
Females (N = 29)		2.0 (1.91, 2.09) mm[b]	
Males (N = 9)			
Van 't Klooster et al.[3]	Very early or advanced hand OA	1.62 ± 0.27 mm[a]	0.99 ± 0.23 mm
Without OA (N = 20)		1.31 ± 0.25 mm[a]	0.71 ± 0.20 mm
With OA (N = 20)			
Angwin et al.[11]	Early RA	1.74 ± 0.019 mm[c]	1.20 ± 0.013 mm
Females (N = 176)		2.03 ± 0 mm[c]	1.41 ± 0.029 mm
Males (N = 69)			
Burghardt et al.[19]	RA	MCP2 (N = 13)	
microCT scanning		1.73 ± 0.25 mm[a]	
RA (N = 16)		MCP3 (N = 13)	
		1.65 ± 0.23 mm[a]	
		MCP4 (N = 12)	
		1.53 ± 0.22 mm[a]	
Mandl et al.[25]	RA	0.3 to 2.7 mm (N = 251)	
(N = 35)			

[a]Mean ± standard deviation
[b]Mean, (95% confidence interval)
[c]Mean ± SEM.

Table 3 Change in JSW in serial images of the same joint with an interval of 6 to 12 months, related to change in Sharp–van der Heijde score (mean ± standard deviation per ΔSHS).

	COBRA (536 pairs)		DREAM (589 pairs)	
ΔSHS_{JSN}	MCP (mm)	PIP (mm)	MCP (mm)	PIP (mm)
0	0.0 ± 0.1 (N = 262)	0.0 ± 0.1 (N = 265)	−0.0 ± 0.1 (N = 289)	−0.0 ± 0.1 (N = 293)
1	(N = 0)	−0.1 ± 0.1 (N = 3)	0.1 ± 0.1 (N = 5)	−0.0 ± 0.0 (N = 2)
2	−0.2 ± 0.2 (N = 5)	−0.1 ± 0.0 (N = 1)		
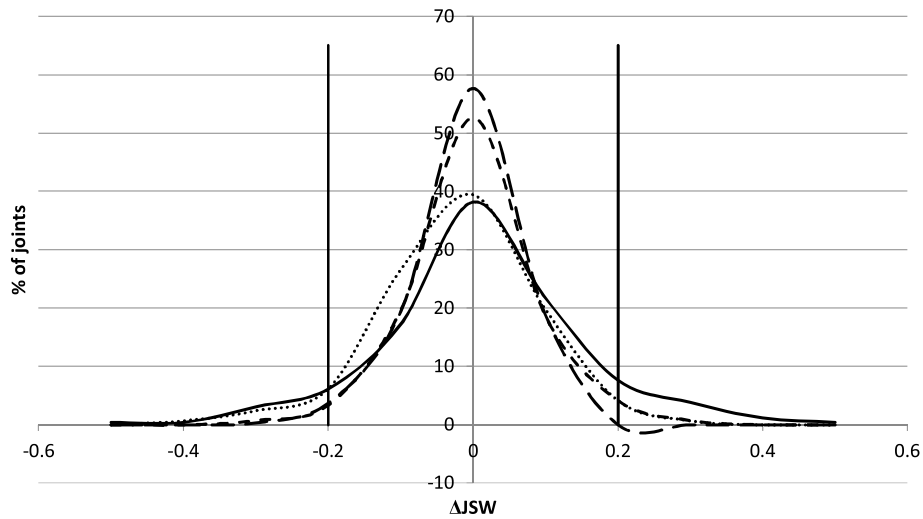
Fig. 2 Graphical representation of the difference in JSW measurements of MCP and PIP joints that was unchanged according to the Sharp–van der Heijde score. Solid line: COBRA MCP, dashed line: COBRA PIP, dotted line: DREAM remission induction MCP, long dash: DREAM remission induction PIP.
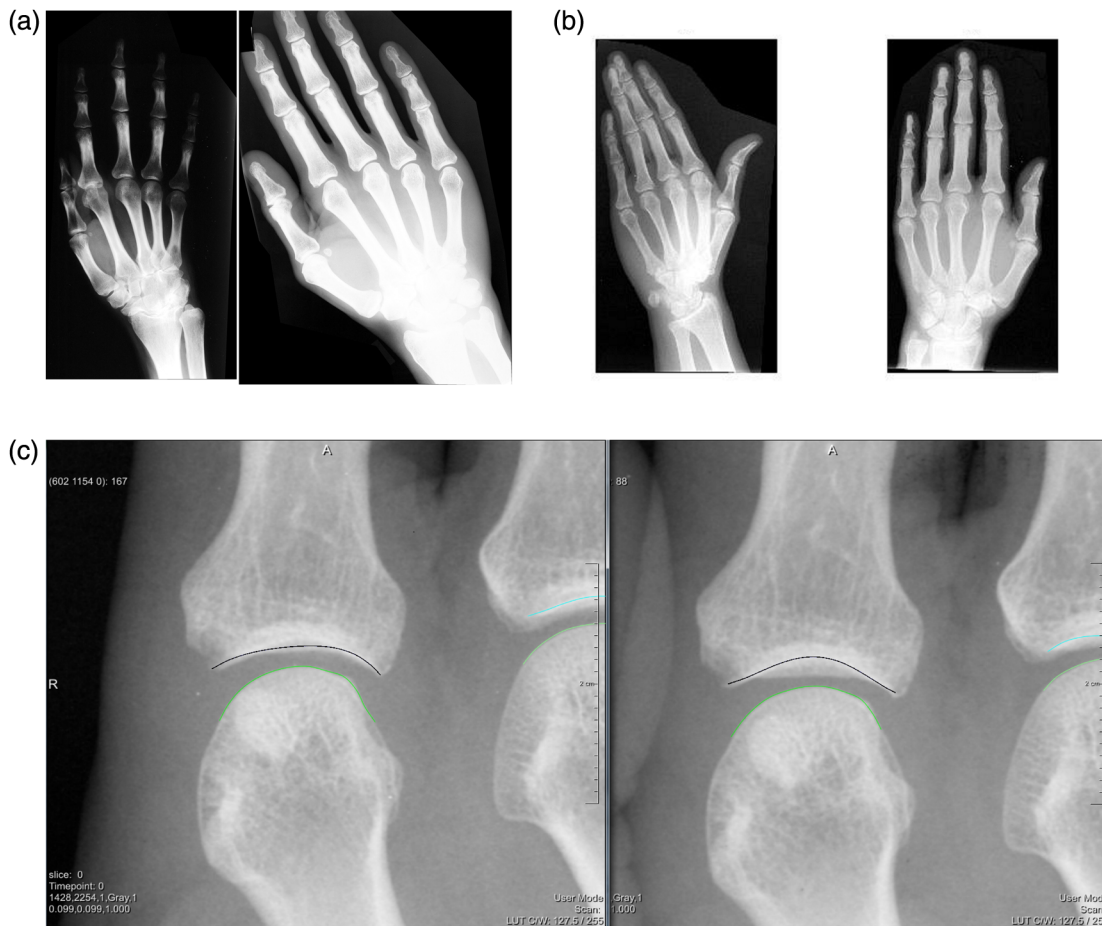
Fig. 3 Examples of errors in image acquisition and margin delineation. (a) Exposure errors: left, too dark; right, too bright; (b) repositioning error: left, t1; right, t2 of the same patient; (c) mismatch in delineation of a joint (left, t1; right, t2 of the same joint).

joints in the COBRA group were scored $\Delta SHS_{JSN} = 2$, and these had mean automated measurements of -0.2 ± 0.2 mm.

Finally, an in-depth review was performed of the 55 joint pairs (COBRA 36, DREAM 19) that fulfilled our predefined criteria for discrepancy between SHS_{JSN} and JSW. The causes for discrepancy as assessed by two rheumatologists were: exposure 15% (COBRA 14%, DREAM 17%), repositioning 57% (COBRA 62%, DREAM 46%), incorrect delineation by the software 25% (COBRA 22%, DREAM 32%), no explanation 3% (COBRA 1%, DREAM 5%). Examples of these causes of error are illustrated in Fig. 3.

6 Discussion

Fully automated measurements of structural joint damage in RA has been a goal for several research groups. Despite promising results, these systems are not used routinely in trials let alone in population wide studies of RA or OA. We believe that structured evaluation will be an important step in acceptance of these methods. In this study, we describe a stepwise validation that may help to further improve automated JSW measurements. The use of images from different sources and the use of consecutive images without change as gold standard demonstrate sources of error and emphasize the importance of a strict protocol when making serial radiographs in the follow-up of RA.

There are no standard methods to validate automated measurements of JSW in plain hand radiographs. Several studies have compared JSW in mm with the current gold standard of joint damage in RA, the SHS. In general, these have produced adequate results, often indicating a somewhat improved sensitivity to change with computer-assisted methods. However, the number of failed measurements and/or the need for human intervention during the "automated" measurements is seldom reported in detail. Dissimilar results in individual joints may not be due to failing software, but can be caused by misinterpretation in the SHS, or variation in image acquisition, such as repositioning of the hand in consecutive images or variations in exposure.

The first step in validation is the success rate of a system to automatically locate joints and joint margins in a large number of images. In the two datasets from different sources in this study, the first step of automated location of the joints was more successful (99% versus 87%) in the DREAM images from a single hospital, than in the multicenter data from the COBRA trial. This underlines the importance of a standard protocol to obtain serial images. We do not believe that the process of digitization of the COBRA images has influenced these results, since the pixel size we used was equal to the DREAM images. Also, we have selected COBRA images that were measured with success in an earlier study[22,23] using an operator mediated system, which biased toward success in these images.

In the second step of automated image analysis, individual joints are measured, which again was slightly more successful (95% versus 92%) in the DREAM images. With operator intervention, the overall performance increased to almost 99%, suggesting that this system provides a rapid tool to measure large numbers of digital radiographs. However, joints that could not be measured because of severe damage need special attention, and future software must be able to assign a correct measure to these in order to gain wider acceptance.

The third step addresses the precision of the measurement result in millimeters. The best validation would be to compare with micro-CT measurements of the 3-D volume of the joint space, but in this study, we have no CT data. The average JSW we found was in line with the results at group level of reported measurements in early RA patients and this provides support for the quality and comparability of the software.

Progression of RA is exclusively measured in serial images over time. Scoring of progression means quantitation of changes between these images. As a fourth step of validation of the automated system, we used pairs of images with little or no change over time as a gold standard. The accuracy of our measurements was good at group level in both datasets, with a mean and standard error of 0.0 ± 0.1 mm in repeated images that are unchanged in the eyes of human observers. This promises sufficient sensitivity to measure changes below the annual joint space decrease of 0.3 mm. As has been pointed out, we may even need higher sensitivity given the improved outcome with current treatment strategies in early RA. In recent years, structural damage has become less common, and therefore, the performance of automated measurements in joints with minimal damage is of prime importance. However, a potential system to measure decreases in JSW will also need tests in images with more advanced damage than we have studied.

In order to improve automated measurements, we must analyze the instances in which the system fails. To this end, we explored in depth the pairs of images that showed discrepancy between the automated measurement and SHS. It turned out that technical factors during the process of image acquisition occurred more frequently than failure of the automated analysis. Differences in positioning and exposure settings caused 9% of the joints to be excluded. Moreover, these influences were the dominant explanation for discrepancy between SHS_{JSN} and automated JSW measurements (repositioning 57% and exposure 15%). Repositioning, as defined by Neumann et al., includes all possible movements of the hand (e.g., flexion, abduction, and adduction), but also radiographic technique, equipment, beam geometry and operator variability. It is considered as a cause or error.[9] Over- or underexposure, which is only detectable after acquisition, was not mentioned by Neumann. All of these factors may increase or decrease the change in JSW in serial images. This is confirmed by our analysis and stresses that standardizing procedures will potentially improve the smallest detectable differences of JSW measurement systems. The importance of quantifying minimal changes or structural integrity was discussed by Landewe and Pfeil.[24,25]

The strength of this validation study is that it is based on a large series of images obtained in different conditions. Furthermore, the use of unchanged joints provides a good reference to validate the accuracy of serial measurements and to explore sources of error. This strength is also a limitation, since we included only a small number of images with serious joint damage. Deformed joints with asymmetric JSN or ankyloses are often hard to measure with fully automated systems. Obviously, it will be important to test systems in groups of patients with longer disease duration. In most studies of automated joint measurements, the lack of measurements on the wrists is another limitation. These are difficult to measure with computerized systems due to the variation in wrist anatomy, repositioning, and the overlap of carpal bones on 2-D images. We may need to explore other approaches than the various efforts to copy the SHS measures of wrist damage. This also raises the question "which joints can reliably and consistently provide information when the goal is to assess long-term

damage in arthritis?"[26] Further research and development of automated assessment of joint damage help to solve these issues.

In conclusion, this study confirms that an automated computerized JSW measurement method for MCP and PIP joints can produce measurements that are in line with the labor intensive SHS_{JSN}. It is obvious that such systems may provide an alternative to human observations and therefore, save time and costs. Standardization of image acquisition is an important and possibly underestimated factor that may help to improve the performance of automated systems. These considerations may support further development of a fully functioning program for quantification of disease progress of RA. This should incorporate not only JSW measurements but also erosion measurements for hands and feet. Accurate measurement of JSW will also be of importance in the follow-up of OA, in particular, in trials of potential interventions for this condition. The proposed method should also be further validated for normal and RA joints in order to establish reference values. In the future, such automated programs may not only be relevant for clinical studies but also be beneficial for assessment of joint damage in clinical practice. This can happen only after careful validation and comparison of these programs.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

References

1. D. van der Heijde et al., "Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis," *Rheumatology* **38**, 941–947 (1999).
2. J. C. Buckland-Wright, I. Carmichael, and S. R. Walker, "Quantitative microfocal radiography accurately detects joint changes in rheumatoid arthritis," *Ann. Rheum. Dis.* **45**, 379–383 (1986).
3. R. van 't Klooster et al., "Automatic quantification of osteoarthritis in hand radiographs: validation of a new method to measure joint space width," *Osteoarthritis Cartilage* **16**, 18–25 (2008).
4. J. T. Sharp et al., "Computer based methods for measurement of joint space width: update of an ongoing OMERACT project," *J. Rheumatol.* **34**, 874–83 (2007).
5. Y. Huo et al., "Automatic joint detection in rheumatoid arthritis hand radiographs," in *Proc. Int. Symp. Biomed. Imaging*, pp. 125–128 (2013).
6. A. Pfeil et al., "Computer-aided joint space analysis (CAJSA) of the proximal-interphalangeal joint-normative age-related and gender specific data," *Acad. Radiol.* **14**, 594–602 (2007).
7. U. Snekhalatha and M. Anburajan, "Computer-based measurements of joint space analysis at metacarpal morphometry in hand radiograph for evaluation of rheumatoid arthritis," *Int. J. Rheum. Dis.* [Epub ahead of print] (2015).
8. J. Böttcher et al., "Computerized quantification of joint space narrowing and periarticular demineralization in patients with rheumatoid arthritis based on digital x-ray radiogrammetry," *Invest. Radiol.* **41**, 36–44 (2006).
9. G. Neumann et al., "Patient repositioning reproducibility of joint space width measurements on hand radiographs," *Arthritis Care Res.* **63**, 203–207 (2011).
10. J. Angwin et al., "Reliability and sensitivity of joint space measurements in hand radiographs using computerized image analysis," *J. Rheumatol.* **28**, 1825–36 (2001).
11. J. C. Buckland-Wright, "Radiographic assessment of osteoarthritis: comparison between existing methodologies," *Osteoarthritis Cartilage* **7**, 430–433 (1999).
12. J. Angwin et al., "Radiographic hand joint space width assessed by computer is a sensitive measure of change in early rheumatoid arthritis," *J. Rheumatol.* **31**, 1050–1061 (2004).
13. A. Finckh et al., "Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: a longitudinal study," *Arthritis Rheumatol.* **54**, 1444–1450 (2006).
14. A. Pfeil et al., "The usefulness of computer-aided joint space analysis in the assessment of rheumatoid arthritis," *Jt. Bone Spine* **80**, 380–385 (2013).
15. C. Lukas et al., "Automated measurement of joint space width in small joints of patients with rheumatoid arthritis," *J. Rheumatol.* **35**, 1288–1293 (2008).
16. Y. Huo et al., "Automatic joint detection in rheumatoid arthritis hand radiographs," in *10th IEEE Int. Symp. Biomed. Imaging*, pp. 125–128 (2013).
17. M. Boers et al., "Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis," *Lancet* **350**, 309–318 (1997).
18. M. Vermeer et al., "Sustained beneficial effects of a protocolized treat-to-target strategy in very early rheumatoid arthritis: three-year results of the Dutch Rheumatoid Arthritis Monitoring remission induction cohort," *Arthritis Care Res.* **65**, 1219–1226 (2013).
19. E. C. Goligher et al., "Radiographic joint space width in the fingers of patients with rheumatoid arthritis of less than one year's duration," *Arthritis Rheum.* **54**, 1440–1443 (2006).
20. A. J. Burghardt et al., "Quantitative in vivo HR-pQCT imaging of 3D wrist and metacarpophalangeal joint space width in rheumatoid arthritis," *Ann. Biomed. Eng.* **41**, 2553–2564 (2013).
21. A. Pfeil et al., "Computer-aided joint space analysis of the metacarpal-phalangeal and proximal-interphalangeal finger joint: normative age-related and gender-specific data," *Skeletal Radiol.* **36**, 853–864 (2007).
22. C. Lukas et al., "Automated measurement of joint space width in small joints of patients with rheumatoid arthritis," *J. Rheumatol.* **35**, 1288–1293 (2008).
23. J. A. Kauffman, C. H. Slump, and H. J. Bernelot Moens, "Detection of joint space narrowing in hand radiographs," *Proc. SPIE* **6144**, 614446 (2006).
24. R. Landewé, V. Strand, and D. van der Heijde, "From inhibition of radiographic progression to maintaining structural integrity: a methodological framework for radiographic progression in rheumatoid arthritis and psoriatic arthritis clinical trials," *Ann. Rheum. Dis.* **72**, 1113–1117 (2013).
25. A. Pfeil et al., "Visualisation of structural damage as a surrogate marker of radiographic progression in patients with rheumatoid arthritis," *Ann. Rheum. Dis.* **73**, e24 (2014).
26. M. A. H. Oude Voshaar et al., "Further simplification of the simplified erosion narrowing score with item response theory methodology," *Arthritis Care Res.* **66**, 1206–1210 (2015).

Olga Schenk finished her bachelor's and master's degree in biomedical engineering at University of Twente in 2013. After that, she started her PhD within the MIRA Research Institute for Biomedical Engineering and Technical Medicine, University of Twente. Her interests are in medical imaging, image analysis, and computer vision. Her current project focuses on improving and evaluating automatic measurement methods of disease progression in rheumatoid arthritis.

Yinghe Huo studied computer engineering, specializing in computer vision, at the Leiden University and graduated in 2010. From 2011, he started as a PhD candidate at the Image Sciences Institute, University Medical Center Utrecht, focusing on automated measurement of joint space width in early rheumatoid arthritis hand radiographs. Since 2016, he has been working at the same faculty, focusing on automatic segmentation of breast cancer and quantification of its response to chemotherapy in contrast-enhanced MRI images.

Koen L. Vincken graduated in informatics (computer science) at the Delft University of Technology in 1989. In 1995, he received his PhD from Utrecht University. He is currently working as an associate professor at the Image Sciences Institute of the Utrecht Medical Center, Utrecht. He developed several applications that are currently being used by other departments, such as radiology, anatomy, orthopedics and rheumatology. His main interests are medical image applications and quantitative image analysis.

Mart A. van de Laar studied medicine at the University of Amsterdam. After training as an internist, he specialized in rheumatology. His expertise lies in the epidemiology of chronic disease especially rheumatic diseases through the use of ICT support; analysis of the perception of health and the development of computer-adaptive testing by making use of modern testing theory.

Ina H. H. Kuper studied medicine at the University of Groningen. She was trained as a rheumatologist and has been practicing rheumatology since 1997 in several hospitals. Her expertise lies in the epidemiology of chronic rheumatic diseases, in particular, rheumatoid arthritis.

Kees C. H. Slump received his MSc degree in electrical engineering from Delft University of Technology, the Netherlands, in 1979. In 1984, he obtained his PhD in physics from the University of Groningen, the Netherlands. From 1983 to 1989, he was employed at Philips Medical Systems in Best, the Netherlands, as head of a predevelopment group on x-ray image quality and cardiovascular image processing. In 1989, he joined the Department of Electrical Engineering, University of Twente. In June 1999, he was appointed as a full professor in signal processing. His research interest is in image analysis as part of medical imaging.

Floris P. J. G. Lafeber is a professor of experimental rheumatology at University Medical Centre Utrecht. He is a manager of research and head of the research laboratory of the Department of Rheumatology & Clinical Immunology, UMCU. Currently, his main interest is in novel treatments; diagnostic tools for automated analyses of radiographic joint damage and inflammation in OA, HA, and RA, as well as unique animal *in vivo* models of osteoarthritis, which have all been developed over the past years.

Hein J. Bernelot Moens is a practicing rheumatologist in large teaching hospitals since 1986 and has been involved in various research projects involving medical imaging of arthritis and its consequences. His research is focused on automated analysis of radiological damage and optical imaging of synovitis.