Title Page

A reliability study using computer-based analysis of finger joint space narrowing in rheumatoid arthritis patients

Katsuya Hatano1, Tamotsu Kamishima2, Kenneth Sutherland3, Masaru Kato4, Ikuma Nakagawa4, Shota Ichikawa1, Keisuke Kawauchi5, Shota Saitou5, Masaya Mukai6

1 Graduate School of Health Sciences, Hokkaido University
North-12 West-5, Kita-ku, Sapporo City, 060 0812 Japan.
2 Faculty of Health Sciences, Hokkaido University
North-12 West-5, Kita-ku, Sapporo City, 060 0812 Japan.
3 Graduate School of Medicine, Hokkaido University
N15 W7, Kita-ku, Sapporo, 060-0815 Japan.
4 Division of Rheumatology, Endocrinology and Nephrology, Hokkaido University Graduate School of Medicine
N15 W7, Kita-ku, Sapporo, 060-0815 Japan.
5 Department of Health Sciences, School of Medicine, Hokkaido University
North-12 West-5, Kita-ku, Sapporo City, 060 0812 Japan.
6 Division of Rheumatology & Hematology, Department of Medicine, Sapporo City General Hospital
1-1, Nishi13-chome, Kita11jo, Chuo-ku, Sapporo-City, 060-8604 Japan

Address correspondence to:
Tamotsu Kamishima, MD, PhD,
Faculty of Health Sciences, Hokkaido University, North-12 West-5, Kita-ku, Sapporo City, 060 0812 Japan.
Phone/FAX; 81-11-706-2824
E-mail: ktamotamo2@yahoo.co.jp

Author identifying information
Conflict of Interest: All the authors declare no conflict of interest.
Funding: This study has no relevant funding.

Abstract

Objectives: The joint space difference index (JSDI) is a newly developed radiographic index which can quantitatively assess joint space narrowing progression of rheumatoid arthritis (RA) patients by using an image subtraction method on a computer. The aim of this study was to investigate the reliability of this method by non-experts utilizing RA image evaluation.

Method: Four non-experts assessed JSDI for radiographic images of 260 metacarpophalangeal (MP) joints from 26 RA patients twice with an interval of more than 2 weeks. Two rheumatologists and 1 radiologist as well as the 4 non-experts examined the joints by using the Sharp-van der Heijde Scoring (SHS) method. The radiologist and 4 non-experts repeated the scoring with an interval of more than 2 weeks. We calculated intra-/inter-observer reliability using the intra-class correlation coefficients (ICC) for JSDI and SHS scoring, respectively.

Results: The intra-/inter-observer reliabilities for the computer-based method were almost perfect (inter-observer ICC, 0.936-0.988; intra-observer ICC, 0.965-0.968). Contrary to this, intra-/inter-observer reliability for SHS by experts was moderate to almost perfect (inter observer ICC, 0.576-0.862; intra observer ICC, 0.654-0.846).

Conclusion: The results suggest that our computer-based method has high reliability to detect finger joint space narrowing progression in RA patients.

Keywords

Rheumatoid arthritis; Joint space narrowing; computer-based analysis; reliability; radiography

Blinded Manuscript

A reliability study using computer-based analysis of finger joint space narrowing in rheumatoid arthritis patients

Introduction

Rheumatoid arthritis (RA) is an autoimmune disease of unknown cause characterized by chronic inflammation of the synovial joints. The resulting joint pain and stiffness may cause impaired function, and for the majority of cases, the progressive synovitis will lead to permanent damage to articular cartilage and bone [1].

Radiograph is the most common modality to assess the degree of the RA progression in daily practice [1]. It is superior to other imaging methods in terms of simplicity, covering a wide range of anatomy, and at relatively low cost and limited harmfulness.

Among methods developed to score the radiographic progression of RA, one of the modifications of the Sharp method, the Sharp/van der Heijde Scoring (SHS) method, is widely accepted in a number of clinical trials [2]. This method has advantages in terms of sensitive scoring for changes in structural damage [1,3]. However, these assessment methods require further study to assess accurately and have a disadvantage in terms of inter- or intra- observer reliability. Sharp et al. reported that the variability in scoring radiograph abnormalities is considerable (inter-reader variability scores assessed by using the intraclass correlation coefficient (ICC) range between 0.465 and 0.999) [4].

Computer-based methods may be more sensitive than traditional methods such as SHS [5–9]. Our study group developed a computer-based quantification of joint space width difference using temporal

subtraction which can detect joint space narrowing (JSN) progression between two sequential radiographic images which we call the joint space difference index (JSDI) [10]. Ichikawa et al. examined JSN progression of 27 RA patients treated with Tocilizumab and disease-modifying antirheumatic drugs (DMARDs) by using a computer-based method. Assessments were carried out by non-experts who had no knowledge such as characteristics of the RA disease or details of the current measurement procedure. They confirmed that when the SHS score increased between baseline and follow-up images, the JSDI tended to demonstrate higher values [11]. However, this computer-based method has not been fully validated in terms of reliability; it includes subjective manual aspects such as ROI setting and image superimposing. Reliability relates the magnitude of the measurement error in observed measurements to the inherent variability in the ‘error-free’, ‘true’, or underlying level of the quantity between subjects [12]. Thus, the purpose of this study was to investigate the reliability of the computer-based method by non-experts by comparing with the SHS method by experts in RA patients.

Materials and Methods

Patients

In this study, 26 RA patients who underwent bilateral hand radiography at baseline and at 1 year

follow-up were included. Average patient age was 58 years old (range; 21-80), and all patients took some kind of medication. Type of medication, average duration of disease and results of pathological examination, such as disease activity score (DAS28), are shown in Table 1 in detail. The study was conducted in accordance with the Declaration of Helsinki [13] and was approved by the local ethics committee. Informed consent was obtained from all patients.

Radiograph acquisition

All radiographs including baseline and follow-up were acquired using X-ray equipment (Toshiba KXO-50G, Tokyo, Japan) under the following routine conditions: no X-ray aluminum filter, film-focus distance of 1 m, tube voltage of 45 kV, tube current of 250 mA, and exposure of 0.014 s. All radiographs were digitized as digital imaging and communication in medicine (DICOM) images with a 0.15 ×0.15 mm pixel size at 10 bit gray-scale.

Joint space narrowing evaluation using Sharp-van der Heijde score

Two rheumatologists and a radiologist who have abundant assessment experience assessed JSN progression using SHS. Their scoring results were treated as the gold standard in this study. The method

of assessment for joint space narrowing was scored as follows: 0 = normal; 1 = focal or doubtful; 2 = generalized, >50% of the original joint space left; 3 = generalized, <50% of original joint space left or subluxation; 4 = bony ankylosis or complete luxation [14].

Computer-based measurement for joint space narrowing progression

The computer-based method can detect the difference in joint space width between the baseline image and the follow-up image with the JSDI. The details of this method are available in a previous paper [10]. Briefly, this method visualizes JSN progression between the baseline image and the follow-up image by displaying narrowing with a red shadow. If there were no changes in joint space width between the baseline and the follow-up images, the joint space in the fused image was visualized as a grey shadow. The JSDI is defined as the average absolute value of the difference of the pixel value in each pixel for baseline and follow-up images inside the region of interest (ROI) (Fig 1).

The computer based method assessed the JSN progression of metacarpophalangeal (MP) joints, using the SHS method assessed by experts as the standard of reference on bilateral hand radiographs. The main aim of the assessment using our computer based method was to detect small joint space narrowing progression at a relatively early joint destruction stage. We thus performed analyses after excluding joints with subluxation or dislocation, determined by a majority vote of experts.

The measurement procedure was performed as follows. First, the software reads the two radiographs for each case and superimposes the baseline image on the follow-up image, assigning cyan to the baseline and red to the follow-up image. Second, the contour of proximal phalanx is superimposed more accurately. Finally, a rectangular ROI with a size of 60 × 20 pixels is set manually in the center of the joint space. The horizontal ROI borders are approximately parallel to the joint edges. The JSDI is then calculated from the base and follow-up pixels within the ROI.

Flow of assessment by SHS and computer based method

Three experts and 4 non-experts performed SHS analysis. One radiologist and 4 non-experts performed SHS analysis twice with an interval of 2 weeks for the purpose of intra-observer reliability. In the same way, 4 non-experts performed computer-based analysis twice with an interval of 2weeks. These intervals were set to assure reproducibility accuracy (Fig 2).

Statistical analysis

Statistical analyses were calculated with the use of SPSS® v. 22.0 (IBM Corp., New York, USA) for Windows® and Excel (Microsoft). Both SHS and computer-based analyses were evaluated for intra- and

inter-observer reliability using the intra-class correlation coefficient (ICC). ICC values were interpreted as follows [15]: the value below 0 was considered poor reliability, slight for values between 0 and 0.20, fair for values between 0.21 and 0.40, moderate for values between 0.41 and 0.60, substantial for values between 0.61 and 0.80, and almost perfect for values between 0.81 and 1.00. -1.0 indicates total disagreement, and 1.0 represents perfect agreement.

Results

The intra-/inter-observer reliabilities using the computer-based method by non-experts was almost perfect (inter-observer ICC for the first session, 0.936; inter-observer ICC for the second session, 0.982; intra-observer ICC, 0.965) (Table 2). After excluding 48 out of a total of 260 joints with advanced JSN / subluxation / dislocation / ankyloses scored by 2 rheumatologists and one radiologist from the reliability assessment, the intra-/inter-observer reliability using the computer-based method was again almost perfect (inter-observer ICC for the first session, 0.942; inter-observer ICC for the second session, 0.988; intra-observer ICC, 0.968) (Table 2). The intra-/inter-observer reliabilities using SHS by experts was also almost perfect (inter-observer ICC for baseline images, 0.846; inter-observer ICC for follow-up images, 0.862, intra-observer ICC for baseline images, 0.846; intra-observer ICC for follow-up images, 0.841) (Table 3). After excluding 48 out of a total of 260 joints with advanced JSN / subluxation /

dislocation / ankyloses scored by 2 rheumatologists and one radiologist from the reliability assessment, the intra-/inter-observer reliabilities using SHS by experts was substantial (inter-observer ICC for baseline images, 0.576; inter-observer ICC for follow-up images, 0.583; intra-observer ICC for baseline images, 0.669; intra-observer ICC for follow-up images, 0.654) (Table 3). Finally, intra-/inter-observer reliabilities for delta values (difference between baseline and follow-up) by experts using SHS were poor to slight (All joints: inter-observer ICC, 0.063, intra-observer ICC, -0.185. After exclusion: inter-observer ICC, 0.137, intra-observer ICC, -0.103).

Discussion

Radiographs are a suitable outcome measure in patients with rheumatoid arthritis [14]. The images reflect the history of the joint pathology and provide a permanent record necessary for serial evaluation of the disease [14]. Use of scored radiographs as an outcome measure can help estimate the progression of rheumatoid arthritis [16]. The most commonly used methods are the Sharp method, one of its modifications being the Sharp–van der Heijde scoring (SHS) method, and the Larsen method with modifications. All these methods require trained readers to obtain sufficient reliability, making them difficult to use in clinical practice. Additional disadvantages for their use in clinical practice include time consumption and cost [16].

We have developed a computer-based method for the assessment of joint space width difference which can objectively detect slight JSN between two images and compute the joint space difference index (JSDI) using a temporal subtraction technique [10,11]. This method allows the objective and quantitative assessment of JSN progression in patients with RA with high sensitivity and specificity [10,11], using a conventional scoring method as the standard of reference. However, the repeatability or intra-/inter-observer reliability of this method has not been fully examined previously.

We therefore assessed the intra-/inter-observer reliability of this computer-based method between non-experts (those who are not rheumatologist or radiologist, and have no experience in reading hand radiographs) in 260 MP joints of 26 RA patients and found intra-/inter-observer reliability was almost perfect. While 2 rheumatologists and 1 radiologist were also invited to compare the reliability of conventional scoring and intra-/inter-observer reliability was also almost perfect. On the other hand, after excluding joints with advanced JSN / subluxation / dislocation / ankyloses, intra-/inter-observer reliability for the computer-based method remained almost perfect but that of the experts was between moderate and substantial. This may mean that the reliability of our computer-based method analyzed by non-experts is equal to or higher than that of conventional scoring by experts.

As was expected, reliability of conventional scoring by experts was higher than that by non-experts. However, reliability of delta values (score difference between baseline and follow-up) for conventional scoring was slight for both between experts and non-experts. Low reliability of delta values for

conventional scoring between experts may be explained by the fact that all of these experts are practicing at different institutions and had little time to practice ahead of the experimental reading. Also, images were provided in chronological order, not in pair for comparison, which is more prone to scoring errors.

Limitations to our study are firstly, reproducibility assessment of JSDI using the conventional scoring method as the standard of reference was not possible due to the small number of joints with JSN progression. Although reproducibility assessment of JSDI was successfully demonstrated in previous papers, we may need to be careful in interpreting the results of this study. Secondly, reliability was investigated on a joint-by-joint basis in this study. Conventional scoring methods provide the score as the summation of each joint score (for example, SHS for JSN ranges from 0 to 120 in the hands) [14], making it difficult to directly compare the reliability. Considering the superior reliability results of experts to those of non-experts, we believe the reliability of the scoring by experts is acceptable.

In conclusion, high reliability of our computer-based method for detecting JSN progression between non-experts was suggested. This result confirms the feasibility of JSN progression analysis by non-experts.

Compliance with Ethical Standards:

Funding: This study has no relevant funding.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

1. van der Heijde DM. Radiographic imaging: the “gold standard” for assessment of disease progression in rheumatoid arthritis. Rheumatology (Oxford). 2000;39 Suppl 1:9–16.

2. Dias EM, Lukas C, Landewé R, Fatenejad S, van der Heijde D. Reliability and sensitivity to change of the Simple Erosion Narrowing Score compared with the Sharp-van der Heijde method for scoring radiographs in rheumatoid arthritis. Ann. Rheum. Dis. [Internet]. 2008;67:375–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17644537

3. van Der Heijde D, Boonen A, Boers M, Kostense P, van Der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. Rheumatology (Oxford). [Internet]. 1999;38:1213–20. Available from:

http://rheumatology.oxfordjournals.org/content/38/10/941.short¥nhttp://rheumatology.oxfordjournals.org/content/38/12/1213.short

4. Sharp JT, Wolfe F, Lassere M, Boers M, Van Der Heijde D, Larsen A, et al. Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. J. Rheumatol. [Internet]. 2004 [cited 2016 Jul 14];31:1062–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15170916

5. James MF, Heald G, Shorter JH, Turner RA. Joint space measurement in hand radiographs using computerized image analysis. Arthritis Rheum. [Internet]. 1995 [cited 2016 Jul 5];38:891–901. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7612038

6. Higgs JB, Smith D, Des Rosier KF, Charlesworth RW. Quantitative measurement of erosion growth and joint space loss in rheumatoid arthritis hand radiographs. J. Rheumatol. [Internet]. 1996 [cited 2016 Jul 5];23:265–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8882030

7. Böttcher J, Pfeil A, Rosholm A, Sörös P, Petrovitch A, Schaefer ML, et al. Computerized quantification of joint space narrowing and periarticular demineralization in patients with rheumatoid arthritis based on digital x-ray radiogrammetry. Invest. Radiol. [Internet]. 2006 [cited 2016 Jul 5];41:36–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16355038

8. Finckh A, de Pablo P, Katz JN, Neumann G, Lu Y, Wolfe F, et al. Performance of an automated computer-based scoring method to assess joint space narrowing in rheumatoid arthritis: A longitudinal

study. Arthritis Rheum. [Internet]. Wiley Subscription Services, Inc., A Wiley Company; 2006 [cited 2016 Jul 4];54:1444–50. Available from: http://doi.wiley.com/10.1002/art.21802

9. Pfeil A, Renz DM, Hansch A, Kainberger F, Lehmann G, Malich A, et al. The usefulness of computer-aided joint space analysis in the assessment of rheumatoid arthritis. Joint. Bone. Spine [Internet]. Elsevier Masson SAS; 2013;80:380–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23245655

10. Ichikawa S, Kamishima T, Sutherland K, Okubo T, Katayama K. Radiographic quantifications of joint space narrowing progression by computer-based approach using temporal subtraction in rheumatoid wrist. Br. J. Radiol. 2015;89.

11. Ichikawa S, Kamishima T, Sutherland K, Okubo T, Katayama K. Performance of computer-based analysis using temporal subtraction to assess joint space narrowing progression in rheumatoid patients. Rheumatol. Int. [Internet]. Springer Berlin Heidelberg; 2016;36:101–8. Available from: "http://dx.doi.org/10.1007/s00296-015-3349-3

12. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. Ultrasound Obstet. Gynecol. 2008;31:466–75.

13. Association WM. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. 1964;1–8.

14. van der Heijde DM. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their

reliability and applicability. Baillieres. Clin. Rheumatol. 1996;10:435–53.

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics [Internet]. 1977 [cited 2016 Jun 30];33:159–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/843571

16. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. Ann. Rheum. Dis. [Internet]. 2001;60:817–27. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1753828&tool=pmcentrez&rendertype=abstract

Tables

Table1. Clinical and laboratory characteristics of patients with rheumatoid arthritis at baseline

	value
total no. of subjects included	26
Age, mean (range) years	58 (25-74)
Sex, female/male	22/4
Duration of symptoms, median (IQR) years	9.93 (1-32)
Follow-up time between first and second radiograph, median (IQR) years	1.09
Structural remission/non remission	17/9
ESR, median (IQR) mm/h	15 (5-60)
CRP level, median (IQR) mg/dl	0.1 (0.0-2.7)
Swollen joint count, median (IQR)	0 (0-2)
Tender joint count, median (IQR)	0 (0-17)
DAS28-CRP, median (IQR)	1.2 (1.0-3.9)
DAS28-ESR, median (IQR)	2.5 (1.1-4.8)
Total SHS for baseline, median (IQR)	36.5 (0-494)
Total SHS for follow-up, median (IQR)	37.5 (0-494)
Biologics, no.	
None	0
Infliximab	10
Adalimumab	4
Tocilizumab	1
Etanercept	10
Abatacept	1

CRP, C-reactive protein; DAS28, disease activity score; ESR, erythrocyte sedimentation rate; SHS, Sharp-van der Heijde score, IQR; interquartile range.

Table 2. Intra-/inter-observer reliability using computer-based method

	subject	ICC (range)	n
Inter-ICC			
1st session	all joints	0.936	4
	after exclusion	0.942	4
2nd session	all joints	0.982	4
	after exclusion	0.988	4
Intra-ICC			
	all joints	0.965 (0.905-0.993)	4
	after exclude	0.968 (0.908-0.995)	4

ICC, intra-class correlation coefficients; n, number of raters.

Range of ICC is not given for Inter-ICC because this statistical analysis for 4 raters was performed simultaneously. Intra-ICCs were given as mean (range) of combination pairs.

Table 3. intra-/inter-observer reliability using SHS

	subject	ICC (range)	n
Inter-ICC			
non-expert			
baseline assessment	all joints (260)	0.718	4
	after exclude (212)	0.471	4
follow-up assessment	all joints	0.701	4
	after exclude	0.465	4
expert			
baseline assessment	all joints	0.846	3
	after exclude	0.576	3
follow-up assessment	all joints	0.862	3
	after exclude	0.583	3
Intra-ICC			
non-expert			
baseline assessment	all joints	0.724 (0.378-0.852)	4
	after exclude	0.495 (0.171-0.706)	4
follow-up assessment	all joints	0.714 (0.334-0.872)	4
	after exclude	0.491 (0.158-0.706)	4
expert			
baseline assessment	all joints	0.846	1
	after exclude	0.669	1
follow-up assessment	all joints	0.841	1
	after exclude	0.654	1

SHS, Sharp-van der Heijde Scoring; ICC, intra-class correlation coefficients; n, the number of raters

Range of ICC is not given for Inter-ICC because this statistical analysis for raters was performed simultaneously.

Intra-ICCs for non-experts were given as mean (range) of combination pairs.

Intra-ICCs for experts were given as value for the pair.

Table 4. intra-/inter-observer delta reliability using SHS

	subject	ICC (range)	n
Inter-ICC			
non-expert			
delta assessment	all joints	0.075	4
	after exclude	0.068	4
Expert			
delta assessment	all joints	0.063	3
	after exclude	0.137	3
Intra-ICC			
non-expert			
delta assessment	all joints	0.144 (0-0.262)	4
	after exclude	0.036 (0.002-0.248)	3*
Expert			
delta assessment	all joints	-0.185	1
	after exclude	-0.103	1

SHS, Sharp-van der Heijde Scoring; ICC, intra-class correlation coefficients; n, the number of raters

Range of ICC is not given for Inter-ICC because this statistical analysis for raters was performed simultaneously.

* ICC cannot be calculated for one rater because the delta values were all zero.

Figure Legend;

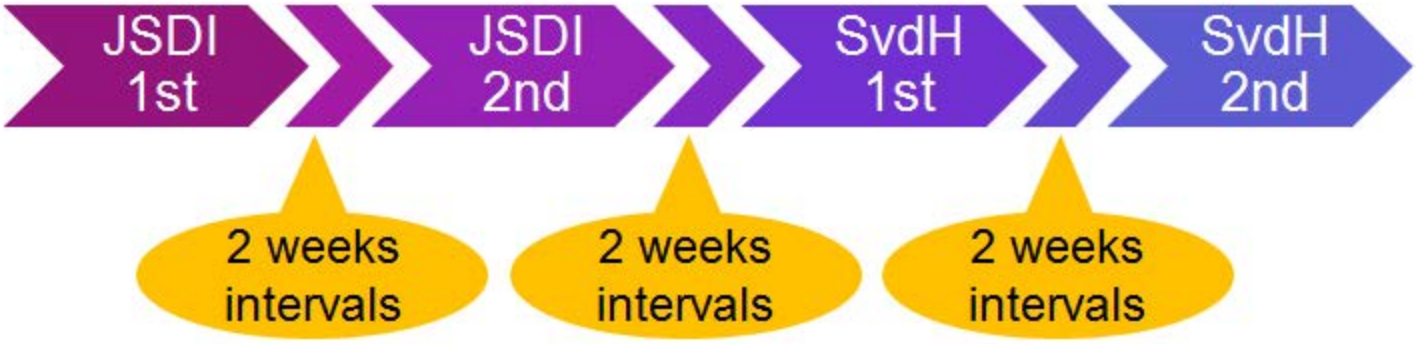
Figure1. ROI setting

A rectangular ROI with a size of 60 × 20 pixels was located manually at the center of the joint space. In this figure, there is no change in joint space width between the baseline and the follow-up images.

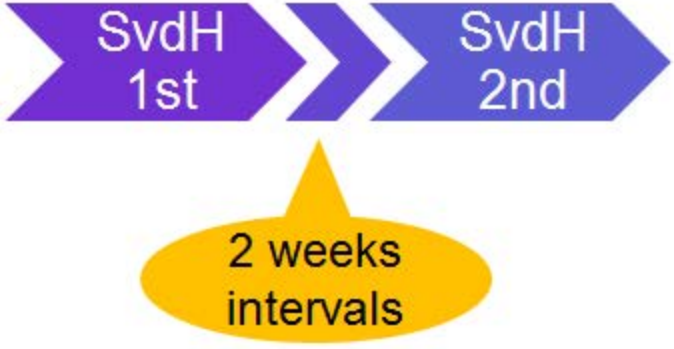
Figure2. Details of assessment flow

There were more than 2 weeks interval between assessments of JSDI $1^{st}/2^{nd}$, JSDI 2^{nd} /SHS 1^{st}, and SHS $1^{st}/2^{nd}$ JSDI, joint space difference index; SHS, Sharp-van der Heijde Scoring.

The flow of assessment by 4 non-experts



The flow of assessment by a radiologist



The flow of assessment by rheumatologists