Scandinavian Journal of Rheumatology

ISSN: 0300-9742 (Print) 1502-7732 (Online) Journal homepage: http://www.tandfonline.com/loi/irhe20

Clinical Reliability of Shoulder Function Assessment in Patients with Rheumatoid Arthritis

C. Boström, K. Harms-Ringdahl & R. Nordemar

To cite this article: C. Boström, K. Harms-Ringdahl & R. Nordemar (1991) Clinical Reliability of Shoulder Function Assessment in Patients with Rheumatoid Arthritis, Scandinavian Journal of Rheumatology, 20:1, 36-48, DOI: 10.3109/03009749109165920

To link to this article: http://dx.doi.org/10.3109/03009749109165920

Scand J Rheumatol 20: 36–48, 1991

Clinical Reliability of Shoulder Function Assessment in Patients with Rheumatoid Arthritis

C. BOSTRÖM, K. HARMS-RINGDAHL and R. NORDEMAR

Department of Rehabilitation & Physical Medicine, Karolinska Institute, Stockholm, Sweden

Boström C, Harms-Ringdahl K, Nordemar R. Clinical Reliability of Shoulder Function Assessment in Patients with Rheumatoid Arthritis. Scand J Rheumatol 1991; 20: 36–48.

A model for functional assessment and a dynamic test of the shoulder joint were designed and tested for normal variation and clinical inter- and intra-rater reliability. The functional assessments, which covered four common shoulder functions, were compared with assessments of pain, recordings of active motion range and the results of a Health Assessment Questionnaire, in eight patients with rheumatoid arthritis according to the ARA criteria. Intra-rater reliability was satisfactory for all four functions and inter-rater reliability was satisfactory for the hand-raising and hand-to-opposite-shoulder functions but less so for hand-behind-back and hand-to-neck. A second test–retest study in 15 patients, with a slight modification of one of the functional tests, confirmed the results and improved the reliability of the modified test. The reliability of the dynamic test and of the active motion range measurement was less satisfactory or not satisfactory.

No significant correlation was found between shoulder functional assessment and the Fries index, but there were positive significant correlations between active motion range and shoulder functions.

It is concluded that the method presented for evaluating shoulder functions has satisfactory reliability and in the first test–retest study was more reliable than conventional motion range measurement of the shoulder joint.

Key words: Shoulder, rheumatoid arthritis, assessment, impairment, disability, evaluation, activities of daily living, function, motion range, pain.

Carina Boström, Dept of Rehabilitation & Physical Medicine, Karolinska Hospital, P.O. Box 60 500, S-104 01 Stockholm, Sweden.

INTRODUCTION

As many as 70–90% (1, 2) of patients with rheumatoid arthritis (RA) have shoulder problems. The resulting disabilities, i.e difficulties in e.g personal hygiene, may be caused by impairment of shoulder joint function.

One goal of physiotherapeutic treatment of the shoulder joint is to reduce disability by restoring impaired joint function. Physiotherapists commonly measure impairment in terms of motion range with a goniometer. However, the reliability varies according to the body region being measured, the movement being measured, whether the measurements are of active or passive motion, the type of goniometer, the patient, the tester, and the clinical situation (3), and the time of day.

The accurate measurement of function ought to be based on several reliable and valid assessments and recordings of joint function (4,5).

Quantification of functional impairment in the shoulder joint guides the examiner in joint function analysis and provides information about the patient's ability to perform certain activities of daily living (ADL). Several authors have reported methods for assessing shoulder function (6, 7, 8, 9, 10,11, 12, 13, 14, 15, 16). The methods are designed mainly for long-term comparisons, for assessments after orthopaedic surgery, as descriptions and patient classifications or they are included in a total functional index.

When considered as clinical tools for treatment evaluation of the shoulder joint, the

methods often include too few functional movements for the shoulder joint and too few points on the scale for discrimination after physiotherapy treatment.

The aim of the present study was to design a model with a scale that is sufficiently detailed for assessing functional impairment in the shoulder joint, and to test the model for normal day-to-day variation and clinical reliability on patients with RA and shoulder problems.

The assessments of shoulder function were also compared with active motion range, assessed pain and the Fries index calculated from the Health Assessment Questionnaire (HAQ) (17), in order to find the normal day-to-day variation and agreement between these variables.

MATERIALS AND METHODS

Subjects in the first test–retest

Eight women with RA according to revised ARA criteria (18) and shoulder problems (pain and/or restricted motion range) were included in the study. The patients were chosen consecutively from among those attending a physiotherapy out-patient clinic at the Karolinska Hospital. Their ages were between 30 and 69 years (mean 51.6 yrs, SD 14.3). Disease duration was between 0.5 and 18 years (mean 7.5 yrs, SD 6.5). The patients belonged to functional classes I–II according to Steinbrocker (19). They were not in an acute phase of the disease and their drug consumption was stable.

The patient's neck problems were not severe enough to warrant X-ray examination. However, all patients had restricted active motion range in the neck; rotation varied from 30° to 70°, lateral flexion from 10° to 40°, flexion/extension from 15° to 25°. The motion range in the elbow was also restricted, mainly as regards extension, which varied between −5° and 35°. The other motion ranges were moderately restricted.

On inquiry the patients mentioned the following disabilities involving the shoulders (one or two examples from each patient): dressing, doing their hair, washing under the arms, pulling up bedspreads, hanging things up and taking things out or down, applying creams and ointments, carrying and lifting things, driving a car.

Subjects in the second test–retest

As the results in the first test–retest study indicated that a slight modification of the hand-to-neck functional test was necessary, a second test–retest study was conducted. Fifteen patients, 11 women and 4 men, with RA according to ARA criteria (18) and shoulder problems (pain and/or restriction in motion range) were included in this study. The patients were chosen consecutively from those attending a Physiotherapy Department at the Karolinska Hospital. Their ages were between 34 and 82 years (mean 62.9, SD 14.8). Disease duration was between 0.5 and 37 years (mean 12.6, SD 10.0). The patients belonged to functional class I–IV according to Steinbrocker (19).

Assessments of shoulder function

A model was constructed for assessing the following four functional shoulder movements:
- hand-raising
- hand-to-opposite-shoulder-behind-back
- hand-to-neck
- hand-to-opposite-shoulder-in-front-of-body.

Each functional movement was assessed on a six-point scale (1 to 6) except for hand-to-neck (seven-points, 1 to 7), giving a total maximum score of 25 for each shoulder. The assessments only took a few minutes per patient. The patient sat on a chair and one shoulder was assessed at a time.

Functional movement scales

I. Hand-raising
Scale scores:
1. The patient does not reach the level of the xiphoid process with the elbow.
2. The patient reaches xiphoid process level with the elbow but with compensatory shoulder elevation.
3. The patient reaches xiphoid process level with the elbow.
4. The patient can raise the elbow to level of the shoulder.
5. The patient can raise the elbow to eye level.
6. The patient can raise the elbow above the head without flexing the neck.

II. Hand-to-opposite-shoulder-behind-back (hand-behind-back)
Scale scores:
1. The patient does not reach the frontal plane through the posterior superior iliac spine (PSIS) with the elbow.
2. The patient reaches PSIS on the ipsilateral side with the dorsal side of the styloid process of the radius.
3. The patient reaches the sacrum at PSIS level with the styloid process of the radius.
4. The patient gets the styloid process of the radius to the contralateral PSIS.
5. The patient gets the styloid process of the radius to the spinous process of the vertebrae at contralateral elbow level.
6. The patient gets the styloid process of the radius past the spinous process of the vertebrae towards the contralateral inferior angle of the scapula.

This function includes shoulder extension, adduction, and internal rotation and elbow mobility.

III. Hand-to-neck
Scale scores:
1. The patient does not reach the fourth cervical spinous process of the vertebrae (C4) with the DIP (distal interphalangeal joint) on the third finger.
2. The patient reaches the cervical spinous process of the vertebrae with the third finger, DIP, but with compensation by flexion and rotation of the neck, adduction of upper arm or shoulder elevation.
3. The patient reaches the cervical spinous process of the vertebrae with the third finger, DIP.
4. The patient reaches the cervical spinous process of vertebrae (C4) with MCP (metacarpophalangeal joint) on the third finger and the elbow reaches shoulder level in the sagittal plane.
5. The patient reaches the cervical spinous process of vertebrae (C4) with the third finger, MCP, and can move the elbow laterally towards the frontal plane but does not reach it.
6. The patient reaches the cervical spinous process of vertebrae (C4) with the third finger, MCP, and can reach the frontal plane with the elbow.
7. The patient passes the cervical spinous process of vertebrae (C4) with the third finger,

MCP, and can move the hand towards contralateral superior angle of scapula with the elbow in the frontal plane.

This function includes shoulder flexion, abduction, external rotation and elbow mobility.

In the second test–retest study, scale score 2 was excluded for assessment of the hand-to-neck movement, as a consequence of the first test–retest study.

IV. Hand-to-opposite-shoulder-in-front-of-body (hand-to-opposite-shoulder)
Scale score:

1. The patient does not reach the contralateral coracoid process of the scapula with MCP on the third finger.
2. The patient reaches the coracoid process of the scapula with the third finger, MCP.
3. The patient reaches around the contralateral spine of the scapula with the third finger, DIP.
4. The patient can reach around the contralateral spine of the scapula with the third finger, DIP, and can lift the elbow to xiphoid process level.
5. The patient can reach around spine of scapula with the third finger, DIP, and can lift the elbow to shoulder level.
6. The patient can reach around spine of scapula with the third finger, DIP, and can lift the elbow to eye level.

This function includes shoulder adduction, flexion, internal rotation, and elbow mobility.

Dynamic test

A dynamic test was also designed for recording the number of hand-raising movements, as far as possible according to the range of motion, that could be performed during 60 seconds.

Active motion range

The active motion ranges of shoulder flexion, abduction, and external rotation vere recorded using a goniometer ad modum Brodin (LIC, Stockholm, Sweden), while the patient sat on a chair. The body segment references for flexion and abduction were the trunk and humerus and for external rotation in the sagittal plane of humerus and ulnae. External rotation was recorded with the arm hanging at the side of the body and the elbow flexed 90°.

Pain assessments

Before the series of shoulder function tests, perceived 'pain-at-rest' was assessed on Borg's category scale from 0 to 10 (20). Pain during each functional test, dynamic test, and active motion range measurement was also assessed. The pain assessments were used to ensure that the pain levels were stable despite all the testing, and as a check on the patients' effort to carry out functional movements and active motions with the same amount of exertion each time.

Fries index

On the first and the third days, the patients answered an HAQ from which a Fries index was calculated (17). This index records the ability to perform ADL and is used for grouping patients into functional classes. The index can vary from 0 to 3: the higher the score, the

more disabled the patient. The Fries index has been translated into Swedish and has good reliability and validity (21).

Measurements of reliability

First test-retest study

Shoulder function was assessed, and active motion range was recorded, for each patient by each of three physiotherapists in random order on each of three days in one week. These tests were included in a routine examination of the shoulder joint and it took one hour per physiotherapist to perform the tests. Each patient was tested at the same time of the day. Thus, both normal variation and inter- and intra-rater reliability could be analysed by calculating agreements or differences. The three physiotherapists had 1–6 years experience of measuring joint function. Before starting, the physiotherapists received verbal and written information about the body segment reference levels to be recorded and the scale scores for the assessments. On each occasion, two trials were performed for each functional movement and for the different motion range directions. The first trial was used for learning and the values from the second were used in the analysis.

Second test-retest study

A second test-retest study was conducted in order to test the reliability of the assessments of the four functions, especially the hand-to-neck function with the exclusion of score 2 from the scale.

Shoulder function was assessed on each patient by each of two physiotherapists in random order on each of two non-consecutive days in one week. In other respects the conditions were as in the first test-retest study. The two physiotherapists had 6 and 11 years experience of measuring joint function.

Statistics

Friedman's two-way analysis of variance by ranks (22) was used to find whether there were any significant changes in assessments of shoulder function, active motion range, and pain with respect to:

- the mean values between occasions per day (i.e systematic changes between the order of assessments per day),
- the mean values between days (i.e systematic changes between days),
- the mean values between physiotherapists (i.e systematic differences between the three physiotherapists).

The Spearman rank correlation coefficient was used to calculate the correlation between functional assessments and (i) active motion range, (ii) HAQ, and (iii) pain.

The Wilcoxon matched pairs signed-ranks test was applied to identify differences in disability Fries index and in shoulder impairment (functional model) between the first test day and the last.

RESULTS

First test-retest study

Shoulder function

The scores included all points on the scale in all functions during the three test days (Fig.1) but levels 4, 5 and 6 were most frequent. The medians were:

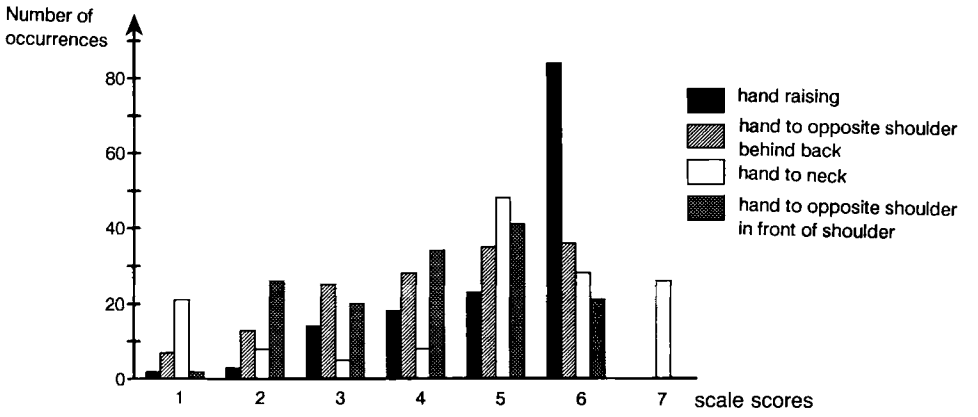

Fig. 1. Occurrence of different scale scores in assessment of the four functions (n=8) (two shoulders/patient).

hand-raising	6 (day 1), 6 (day 2), 6 (day 3)
hand-behind-back	4 (day 1), 5 (day 2), 4 (day 3)
hand-to-neck	5 (day 1), 5 (day 2), 5 (day 3)
hand-to opposite shoulder	4 (day 1), 4 (day 2), 4 (day 3)

The total scores (two shoulders/patient, n=8) for the four shoulder-functional movements assessed by one physiotherapist, on days 1 and 3 respectively, were calculated. Median total score for the eight patients on day one was 38 points (range 17–45) and on day three, 39 points (19–43). There were no significant differences here, nor were there any differences between the other two physiotherapists' assessments.

Functional movement test

There was no significant difference between the order of assessments on each day or between days. However, there was a significant difference between physiotherapists in the assessments of hand-behind-back (left shoulder) ($p<0.05$) and in hand-to-opposite-shoulder (right shoulder) ($p<0.05$)

Inter-rater reliability calculated as percentage agreement per day on all assessments for each functional movement is presented in Table I. Agreement was considered to obtain either when the three physiotherapists' assessments varied by one point on the scale or when they were identical. When one scale point of difference was allowed, agreement for the four shoulder functions varied from 54 to 81%. When it was required that all three assessments be identical, agreement varied from 13 to 51%.

The differences between the highest and lowest assessed values per patient and function were calculated. The means of the differences varied between 1.9 and 0.9 scale points for the first day and 1.1. and 0.6 for the third day.

The intra-rater reliability for each functional movement is also presented in Table I. When one point of difference was allowed, agreement varied from 75 to 87% and when it was required that all three assessments be identical, agreement varied from 29 to 64%.

The differences between the highest and lowest values assessed per physiotherapist and function were calculated. The means of the differences varied over the three days between 0.4 and 1.6 scale points.

Table I. *Inter- and intra-rater reliability for each function*

	Agreement within same scale score (%)		Difference of one scale point (%)	
	Inter	*Intra*	*Inter*	*Intra*
Hand-raising	51	64	81	85
Hand-behind-back	13	33	65	75
Hand-to-neck	19	40	54	87
Hand-to-opposite-shoulder	17	29	81	85

Table II. *Mean values for active motion range day 1 and day 3.*

	Day 1		Day 3	
	$\bar{X}$	(SD)	$\bar{X}$	(SD)
Flexion	119°	(27.6)	125°	(22.6)
Abduction	83°	(31.0)	77°	(22.0)
External rotation	36°	(10.5)	35°	(10.8)

Table III. *Inter- and intra-rater reliability for each active motion range*

	Agreement within 5° (%)		Agreement within 10° (%)	
	Inter	*Intra*	*Inter*	*Intra*
Flexion	17	23	31	52
Abduction	2	15	19	33
External rotation	21	21	52	56

Dynamic test

The mean value for the numbers of hand-raising movements/60 seconds on day 1 was 24.9 (SD 7.8, range 13–45). On day 3 the mean value was 28.5 (SD 8.4, range 19–50). The numbers increased significantly between days (left shoulder) ($p<0.05$) and between occasions/days (right shoulder) ($p<0.05$), but there were no significant differences between the physiotherapists' assessments.

Inter- and intra-rater reliability were calculated as the percentage agreement of all assessments for the dynamic test. Agreement was considered to obtain when all three physiotherapists' assessments were within five numbers of the hand-raising movement. Agreement was 79% for inter-rater reliability and 46% for intra-rater reliability.

The differences between highest and lowest values per patient were calculated. For inter-rater reliability the means of the differences varied between 5.7 for the first day and 3.1 for the third day. For intra-rater reliability the means (based on the differences between highest and lowest values per physiotherapist and function) varied between 6.3 and 6.9 over the three days.

Table IV. *Assessed pain (Borg scale 0–10) (medians) days 1, 2 and 3 during each functional test, dynamic test and active motion range*

Pain	Day 1 median	Day 2 median	Day 3 median	Range day 1–3
Hand-raising	2	1	1	0–7
Hand-behind-back	3	3	3	0–8
Hand-to-neck	2	2	3	0–7
Hand-to-opposite-shoulder	3	2	3	0–10
Dynamic test	2	2	2	0–7
Active flexion	1.5	2	1	0–5
Active abduction	3	3.5	3	0–7
Active external rotation	2	2	2	0–7

Table V. *Correlation coefficient (r) between assessment of shoulder function and active motion ranges. Means of all assessments and measurements over the three days*

	Active flexion		Active abduction		Active ext. rotation	
	right	left	right	left	right	left
Hand-raising	.81*	.92**	.78*	.66	−.49	−.61
Hand-behind-back	.74*	.21	.67	.52	−.09	−.48
Hand-to-neck	.74*	.72	.47	.25	.00	.14
Hand-to-opposite-shoulder	.86*	.80*	.69	.32	.21	.24

*Significance $0.01 < p < 0.05$
**Significance $0.05 < p < 0.001$

Table VI. *Inter- and intra-rater reliability for each function. Second test-retest study*

	Agreement within same scale score (%)		Difference of one scale point (%)	
	Inter	*Intra*	*Inter*	*Intra*
Hand-raising	55	62	82	90
Hand-behind-back	57	75	95	98
Hand-to-neck	63	53	80	73
Hand-to-opposite-shoulder	63	65	88	92

Active motion range

The mean values for the active motion ranges on day 1 and day 3 are presented in Table II.

Intra- and inter-rater reliability are presented in Table III.Assessments within 10° or 5°, respectively, were considered to agree. When a 10° difference was accepted, agreement varied from 33 to 56% for intra-rater reliability and from 19 to 52% for inter-rater reliability. When a 5° difference was accepted, agreement varied from 15 to 23% for intra-rater reliability and from 2 to 21% for inter-rater reliability.

The differences between the highest and lowest values per patient and motion range were calculated. For inter-rater reliability the means varied between 11° and 33° on the first day

and 7° and 19° on the third day. For intra-rater reliability, the mean differences (based on the differences between the highest and lowest values per physiotherapist and motion range) varied from 6° to 28° over the three days.

There were significant differences between physiotherapists for flexion (left shoulder) ($p<0.05$), abduction (right shoulder) ($p<0.001$), and external rotation (right shoulder) ($p<0.05$). There were significant differences between occasions for abduction (right, left shoulder) ($p<0.05$) and external rotation (right shoulder) ($p<0.05$). There were significant differences between days for flexion (right shoulder) ($p<0.05$).

Pain

Median perceived pain-at-rest before performing the functional movements was '0 – no pain' on both days 1 and 3 (ranges 0–4 and 0–3 respectively). Thus, there was no significant change in perceived pain-at-rest during the week. Median pain during each functional test and active motion range are presented in Table IV.

There was no significant difference between perceived pain for (i) different occasions/day, (ii) different days, or (iii) different physiotherapists during the assessments of the functional movements and measurements of active motion range.

Fries index

The median Fries index on the first test day was 1.30 (0.38–1.75) and on the last day 1.00 (0.63–1.75) scores. There was no significant change between the first and the third test days.

Shoulder-functional movements – active motion range

The active motion ranges were compared with the assessments of shoulder function for each patient. The four functional movements all correlated significantly with active flexion, (right shoulder) ($p<0.05$) (Table V). Hand-raising and hand-to-opposite-shoulder also correlated with flexion, (left shoulder) ($p<0.01$, $p<0.05$) and hand-raising (right shoulder) correlated with abduction (right shoulder) ($p<0.05$) (Table V).

Shoulder function – pain

There was a negative (i.e. less pain – better function) significant correlation between the hand-behind-back functional movement (right side) ($r=0.76$, $p<0.05$) and pain during movement. No other correlations were found.

Shoulder function – the Fries index

No significant correlation was found between single shoulder functions (right, left shoulder) and the Fries index.

Second test-retest study

Shoulder function

The scores included all points on the scale in all functions during the two test days. The medians were:

hand-raising	4 (day 1), 5 (day 2) range 1–6
hand-behind-back	5 (day 1), 5 (day 2) range 1–6

hand-to-neck	3 (day 1), 3 (day 2) range 1–6
hand-to-opposite-shoulder	4 (day 1), 4 (day 2) range 1–6

Functional movement test

Inter-rater reliability calculated as percentage agreement per day on all assessments for each functional movement is presented in Table VI. When one scale point of difference was allowed, agreement for the four shoulder functions varied from 80 to 95%. When it was required that both assessments be identical, agreement varied from 55 to 63%.

Intra-rater reliability for each functional movement is also presented in Table VI. When a difference of one point was allowed, agreement varied from 73 to 98%. When it was required that both assessments be identical, agreement varied from 53 to 75%.

DISCUSSION

The clinical intra-rater reliability of assessments of shoulder functional movement was satisfactory, as 75–87% of the assessments differed by not more than one scale point in the first study and 73–98% in the second study. Also, the clinical inter-rater reliability for the functions hand-raising and hand-to-opposite-shoulder was satisfactory in the first test-retest study, with 81% of the assessments within one scale point of difference. However, the assessments of the functions hand-behind-back and hand-to-neck were less satisfactory, with 54–65% of the assessments within one scale point of difference. In the second test-retest study the inter-rater reliability was satisfactory, with 80–95% of the assessments within one scale point of difference. Similar conclusions about reliability have been drawn by Boeckstyns et al (23), who stated that when 84% of the assessments on a visual analogue scale (0–10 levels) differed by not more than one point, the reliability was highly satisfactory. These authors observe that perfect agreement between evaluations will hardly ever occur because of pain, etc. Since RA patients' symptoms vary over time because of e.g. disease activity and pain, a certain inter- and intra-rater difference over time can be expected.

As expected, intra-rater reliability was higher than inter-rater reliability for the functional movement assessment, which is in accordance with the results of other studies on motion range measurements (24). The reliability in our study would have been better with fewer points on the scale. In a study (16) using three scores for functional assessment of the hand and the limbs, the reliability was considered satisfactory, with the inter- and intra-observer correlation coefficient varying between 0.55 and 0.96. The purpose of these functional assessments (16) was to detect early manifestations of RA whereas our aim was to find the number of scale points that would provide enough discrimination for treatment evaluation. However, the difference in functional performance between the different scale scores is small. The results of the inter- and intra-rater reliability tests illustrate this, as the differences between agreement within the same scale score and with one point of difference are apparent. However, further studies are in progress to evaluate the sensitivity.

The inter-rater reliability for the functions hand-to-neck and hand-behind-back was less satisfactory in the first study. This might be because the movements are complicated to assess and include components of rotation. Also, in the scaling for the hand-to-neck function, errors in assessment at 2 point on the scale may distort the results for the others because this score assesses mechanisms of shoulder and neck compensation in the effort to reach the neck. In the second test-retest study, when point 2 was excluded in the assessment of the hand-to-neck function, the inter-rater reliability improved. The reliability also improved for all functions, especially for the function hand-behind-back and hand-to-opposite-shoulder. The reason for this could be that there were only two assessors and two occasions

in the second study. In addition, the physiotherapists in the second study had more clinical experience in measuring joint function and assessing shoulder-function movements according to the present model than the physiotherapists in the first.

For ethical reasons, since all tests in the first study were performed on each patient, it was necessary to choose functional class II patients without great disability. The patients' pain values were low. Consequently, although functional movement was assessed at all points on the scale, 4, 5 and 6 were the most frequent scores.

The patients dynamic test performance improved significantly from occasion to occasion and from day to day. The possible therapeutic effect of the movements on which the successive measurements were based might explain the significant changes. Another reason might be that the patients remembered their scores and tried to improve. The reliability of the dynamic test was less satisfactory or not satisfactory, since 79% (inter-rater reliability) and 46% (intra-rater reliability) of the assessments were in agreement within five numbers. As the dynamic test is influenced by other factors (e.g. motivation, pain) as well as by shoulder involvement, this test cannot be recommended as a test of shoulder fatique.

Recorded active abduction and external rotation differed significantly between occasions/day and between physiotherapists. The physiotherapists probably differed in the amount of attention they paid to whether or not, the shoulder was raised in the measurement of abduction.

At 10° difference, the clinical intra- and inter-rater reliability for active motion range was considered less satisfactory, and at 5° it was considered not satisfactory. Boone et al (24) have suggested that »when more than one tester measures the same motion, changes in range of motion should exceed 5° for upper extremity motion in order to be able to state that improvement has occurred«. How much of a difference should be accepted as normal variation is a matter for discussion. It may be that the motion range of RA patients varies more than 10° over a test period of three days and/or that the physiotherapists differed in their standardization procedures such as use of body segment references. Thorough discussions between clinicians about measurement standards seem to be necessary.

Neither the total sum of the disability index Fries index nor the total sum of the shoulder-function movement assessments changed significantly from day 1 to day 3. There were no significant correlations between the values of any single shoulder function and the total disability index; not surprisingly, since Fries index does not measure small changes in impairment level, only activities at disability levels involving more than one joint. McCloy et al (25) found low positive correlation between dressing/undressing and range of shoulder external rotation, and concluded that high correlations are not to be expected, because an individual's functional ability is influenced by a complex interaction of physical, environmental, and physiological variables. However, Badley (26) has shown that shoulder and arm involvement in RA is reflected in the way difficulties are encountered when, for example having a bath or an all-over wash, dressing, doing one's hair, and eating. Eberhardt et al (16) found significant correlations between Fries index and functional measurements for use in early screening of manifestations of RA. Lower-limb items correlated better than upper-limb items. The assessments in these authors' study resemble the assessments described in the present study and one can conclude that the relations between impairment in different upper-extremity joints and disability must be studied further.

The reliability of the measurements of active motion range was less satisfactory or not satisfactory, which affects conclusions about validity. However, there were positive significant correlations for both shoulders between active flexion and hand-raising and between active flexion and hand-to-opposite-shoulder. In an earlier study (27), a correlation between shoulder abduction and reaching above the head was found. In the present study such correlation was found only for the right shoulder.

The pain reported during the functional movements was fairly low, which can explain why only one correlation between pain and performance of functional movements was found. Perceived pain at rest and during the different movements did not change significantly between days, occasions per day or physiotherapists. Thus, there was no sign that the tests had increased the pain at rest. In addition, the patients appeared to always exert the same amount of effort in performing the function.

It is concluded that the method presented for evaluating shoulder functions has satisfactory clinical reliability if the results from both test-retest studies are taken into account. However, the normal day-to-day variation in motion range seems large, which creates a problem for treatment evaluations, and base-line recordings seem indicated. In the first test-retest study the assessment of shoulder function were more reliable than conventional active motion range measurement of the shoulder joint.

ACKNOWLEDGEMENTS

Grants were received from the Karolinska Institute and the Clas Groschinsky Foundation, Sweden.

The Swedish Association against Rheumatism and the Swedish Medical Research Counsil B 89–19X–05720–10C, B 90–19X–05720–11A are thanked for their support.

REFERENCES

1. Pahle JA, Kvarnes L. Shoulder synovectomy. Ann Chir Gynecol 1985; 74(Suppl 198): 37–9.
2. Petersson CJ. Painful shoulders in patients with rheumatoid arthritis. Prevalence, clinical and radiological features. Scand J Rheumatol 1986; 15: 275–9.
3. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther 1987; 67: 1867–72.
4. Granger CV, Gresham GE. Functional Assessment in Rehabilitation Medicine. Baltimore/London: Williams Wilkins, 1984; 343–63.
5. Jette AM. Functional Status Index: Reliability of a chronic disease evaluation instrument. Arch Phys Med Rehabil 1980; 61: 395–401.
6. Eberl DR, Fasching V, Rahlfs V, Schleyer I, Wolf R. Repeatability and objectivity of various measurements in rheumatoid arthritis. A comparative study. Arthritis Rheum 1976; 19: 1278–86.
7. Wingate L. Efficacy of physical therapy for patients who have undergone mastectomies. Phys Ther 1985; 65: 896–900.
8. Granger CV. Medical Rehabilitation Research and Training Center. Annual Progress Report 1974; no 7: Boston: Tufts University School of Medicine.
9. Jones E, McNitt B, McKnight E. Patient Classification for Long-Term Care: User's Manual. DHEW Publication 1974; No HRA 75–3107: Washington D.C.: U.S. Government Printing Office.
10. Ehrlich G. Assessment of Function: Total Management of the Arthritic Patient. Philadelphia: JB Lippincott CO, 1973; 233–9.
11. Wolfgang GL. Surgical repair of tears of the rotator cuff of the shoulder. Factors influencing the results. Bone Joint Surg 1974; 56 A: 14–26.
12. Lundgren-Lindquist B, Sperling L. Functional studies in 79-year-olds. II. Upper extremity function. Scand J Rehab Med 1983; 15: 117–23.
13. Solem-Bertoft E, Lundh I, Ringqvist I. Physiotherapy after fracture of the proximal end of the humerus. Scand J Rehab Med 1984; 16: 11–6.
14. Rasmussen JO, Horslev-Petersen K, Helin P. Funktionsbevarende kontrolsystem ved reumatoid arthritis. Gråsten: Kong Christian Den Tiendes Gigthospital, 1984 (in danish).
15. Constant CR, Murley AHG. A clinical method of functional assessment of the shoulder. Clinical Orthopaedics and Related Research 1987; Number 214: 160–4.
16. Eberhardt KB, Svensson B, Moritz U. Functional assessment of early rheumatoid arthritis. Br J Rheumatol 1988; 27: 364–71.
17. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980; 23: 137–45.
18. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, Healey LA, Kaplan SR, Liang MH, Luthra HS, Medsger, Jr TA, Mitchell DM, Neustadt DH, Pinals RS, Schaller JG, Sharp

JT, Wilder RL, Hunder GG. The American Rheumatism Association 1987. Revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 1988; 31: 315–24.
19. Steinbrocker O, Traeger CH, Batterman, RC. Therapeutic criteria in rheumatoid arthritis. JAMA 1949; 140: 659.
20. Borg G. A category scale with ratio properties for intermodal and interindividual comparisons. In Proceedings 1982 of the XXIInd Int. Congr. of Psychology 1980; Leipzig, GDR.
21. Ekdahl C, Eberhardt K, Andersson SI, Svensson B. Assessing disability in patients with rheumatoid arthritis: use of a Swedish version of the Stanford Health Assessment Questionnaire. Scand J Rheumatol 1988; 17: 263–71.
22. Siegel S. Nonparametric Statistics for the Behavioral Sciences. London: McGraw-Hill International Book Company, 1956; Chap 7: London: 166–72.
23. Boeckstyns MEH, Backer M. Reliability and validity of the evaluation of pain in patients with total knee replacement. Pain 1989; 38: 29–33.
24. Boone DC, Azen SP, Lin CM, Spence C, Baron C, Lee L. Reliability of goniometric measurements. Phys Ther 1978; 58 11: 1355–60.
25. McCloy L, Jongbloed L. Robinson Bashall functional assessment for arthritis patients: reliability and validity. Arch Phys Med Rehabil 1987; 68: 486–9.
26. Badley EM, Lee J. Impairment, disability and the ICIDH model III: Underlying disease, impairment and disability. Int Rehabil Med 1987; 8: 174–81.
27. Badley EM, Wagstaff S, Wood PHN. Measures of functional ability (disability) in arthritis in relation to impairment of range of joint movement. Ann Rheum Dis. 1984; 43: 563–9.