Postprint Version	1.0
Journal website	http://linkinghub.elsevier.com/retrieve/pii/S0003999305001978
Pubmed link	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=16003664&query_hl=15
DOI	10.1016/j.apmr.2004.12.031

From the Institute for Research in Extramural Medicine (Terwee, de Winter, Scholten, Jans, Devillé, Bouter) and Department of Rheumatology (van Schaardenburg), VU University Medical Center, Amsterdam, the Netherlands.

Presented in part to the American College of Rheumatology and the Association of Rheumatology Health Professionals, November 1999, Boston, MA.

Supported by the Foundation Stichting Anna Fonds.

Interobserver Reproducibility of the Visual Estimation of Range of Motion of the Shoulder

CAROLINE B. TERWEE, PHD, ANDREA F. DE WINTER, PT, PHD, ROB J. SCHOLTEN, MD, PHD, MARIELLE P. JANS, PT, PHD, WALTER DEVILLÉ, MD, PHD, DIRKJAN VAN SCHAARDENBURG, MD, PHD, LEX M. BOUTER, PHD

ABSTRACT.

Objectives: To assess interobserver reproducibility (agreement and reliability) of visually estimated shoulder range of motion (ROM) and to study the influence of clinical characteristics on the reproducibility.

Design: Test-retest analyses.

Setting: Various health care settings in the Netherlands.

Participants: Consecutive patients with shoulder complaints (N=201) referred by 20 general practitioners, 2 orthopedic physicians, and 20 rheumatologists.

Interventions: Not applicable.

Main Outcome Measures: Independent visual estimation by 2 physiotherapists of the ROM. Agreement was calculated as the mean difference in visual estimation between examiners ±1.96 × standard deviations of this mean difference. The intraclass correlation coefficient (ICC) was calculated as a measure of reliability, based on a 2-way random effects analysis of variance.

Results: The lowest level of agreement was for visual estimation of active and passive elevation (limits of agreement, -43.4 to 39.8 and -46.7 to 41.5, respectively, for the difference between the affected and contralateral sides), for which the level of agreement was most clearly associated with pain severity and disability. The ability to differentiate between subjects was acceptable for all movements for the difference between the affected and contralateral sides (ICCs, >.70) except for horizontal adduction (ICC=.49).

Conclusions: Interobserver agreement was low for the assessment of active and passive elevation, especially for patients with a high pain severity and disability. Except for horizontal adduction, visual estimation seems suitable for distinguishing differences between affected and contralateral ROM between subjects.

COMPLAINTS OF SHOULDER DYSFUNCTION are often associated with restricted range of motion (ROM).[1-4] The difference between the affected and the contralateral sides is related to the

ability to perform the activities of daily life, such as dressing, washing, and lifting above shoulder level.[4-6] Measurement of the ROM plays a vital role in diagnosis,[7-9] assessment of the severity of disability, and the assessment of treatment outcome in clinical practice and research.[9]

Several methods are available for quantifying ROM, such as goniometry, inclinometry, and even high-speed cinematography. High reproducibility is an important prerequisite if measurements are to be useful for discriminative and evaluative purposes. The reproducibility of the various devices used to assess shoulder ROM has been evaluated in several studies.[10-19] Although in clinical practice ROM assessment is often based on visual inspection of the movement, insight into the reproducibility of the visual estimation of shoulder ROM is limited.[20,21]

The level of reproducibility of ROM measurement may be influenced by many factors such as the instruments and procedures applied, the joint examined, or the type of movement tested.[22] Recently, we[23] showed that the interobserver agreement on the diagnostic classification of shoulder disorders was associated with severe pain, chronic complaints, and bilateral involvement. It is unclear whether the reproducibility of ROM measurements might also be influenced by clinical characteristics of the patients.

Our first objective in this study was to evaluate, in a large population of patients, interobserver reproducibility (agreement and reliability) with regard to active elevation and several passive shoulder movements. Our second objective was to evaluate whether clinical characteristics are associated with the level of interobserver reproducibility.

METHODS

Participants

Consecutive eligible patients with shoulder complaints were invited to participate in this study by 20 general practitioners, 2 physicians working in an orthopedic practice, and 20 secondary care rheumatologists. Patients were eligible for participation if they gave informed consent, were between 18 and 75 years of age, and were able to complete questionnaires (eg, no dementia). Patients with shoulder problems resulting from neurologic, vascular, or internal disorders; systemic rheumatic diseases; fractures; or dislocations were excluded. The study was approved by the local institutional review board of the VU University Medical Center.

Design

Within the framework of a study on interobserver agreement of the diagnosis of shoulder disorders, which involved history taking and physical examination,[23] we evaluated the interobserver agreement on visual estimation of ROM of the shoulder joint. After history taking, 2 examiners (MPJ, AFW), both physiotherapists with 3 and 10 years of clinical experience, respectively, independently assessed each patient's shoulder ROM. In their work as physiotherapists, both examiners use visual inspection on a regular basis to diagnose and treat musculoskeletal problems. A second physical examination was performed within 1 hour after the first examination. To prevent the occurrence of systematic differences between examiners because of repeated testing, the sequence of the examiners was randomly allocated. During the preparatory phase of the project, we standardized all measurements to prevent large ranges in the estimation of the ROM due to large differences in the physical examination. Furthermore, during this phase we also investigated whether the physiotherapists assessed the ROM with the patients' arms in similar fixed positions or whether there were different interpretations between the physiotherapists. To verify the visual estimation of ROM, different joint positions were assessed visually and compared with goniometric measurement in the training phase.

Visual Estimation of ROM

Each examiner made 1 assessment of the ROM of both shoulders of each patient. In 1990, the Dutch College of General Practitioners developed guidelines for diagnosis and management of shoulder pain,[24] which are largely based on the concepts of Cyriax.[8] The movements in this study are the movements described in these guidelines.

Patients were seated during all tested movements. Their arms were stretched alongside the body with the thumbs directed forward; this was considered the starting position of 0° for the bilateral elevation. Each patient was asked to raise the stretched arms sideways in the frontal or coronal plane

as far as possible, which is referred to as *active elevation*. ROM of the active elevation ended when, according to the patient, further movement was impossible because of weakness or pain, or when the maximum ROM had been achieved. Then the examiner performed the same movement for each arm, during which the patient was asked to relax the arm; this was called *passive elevation*. The examiner moved the arm until pain limited the ROM or the maximum range (opportunity to note the end-feel) was reached. To assess the amount of passive scapulohumeral abduction, the examiner performed the same movement while palpating the lower angle of the scapula with the thumb, until the scapula began to rotate or pain limited any further motion. Passive scapulohumeral abduction was defined as the amount of movement existing between the scapula and the humerus during this movement.[8]

The starting position for assessment of the passive external rotation was with the upper arm in 0° elevation, the elbow flexed, and the thumb up. With the elbow held in 90° flexion against the patient's side, the examiner moved the arm in the transverse plane as far outward as possible while the patient was asked to relax. The examiner moved the patient's arm until pain limited the ROM or the maximum range (opportunity to note the end-feel) was reached.

With the arm straight, the arm was moved upward in the sagittal plane until 90° of anteflexion (arm straight forward) to reach the starting position of 0° for the evaluation of the range of passive horizontal adduction. Then the examiner moved the arm in the transverse plane toward the midline of the body as far as possible, while the patient was asked to relax that arm. Again, the examiner moved the arm until pain limited the ROM or the maximum range was reached.

All ROM measurements were recorded in degrees (with increments of 5°).

Additional Assessments

During history taking, examiners recorded demographic characteristics (age, sex) and clinical characteristics (eg, duration of the shoulder complaints, bilateral involvement, pain at rest, sleep disturbances, aggravating factors).

After the physical examination, both examiners independently estimated the severity of pain on a 100-mm visual analog scale (VAS) ranging from 0 (no pain) to 100 (very severe pain). Pain severity was also recorded when the physical examination of the shoulder was hampered by pain or insufficient relaxation.

Before the diagnostic procedure was started, the participants completed several questionnaires. Examiners were blinded to the results, because the subjects' answers might have influenced their assessments of the shoulder complaints. All patients recorded the severity of their pain in the past 24 hours and in the preceding week, at night and during the day, on a VAS with a range as specified above. They also completed the Shoulder Disability Questionnaire (SDQ), which consists of 16 questions pertaining to difficulties in performing various daily activities on the previous day.[25,26] The total score ranges from 0 (no disability) to 100 (difficulty with all applicable items).

Statistical Methods

For each examiner, the mean, standard deviation (SD), and range were calculated for each ROM assessed and for the difference between both sides to show the difference between the affected and the contralateral sides. In clinical practice the difference between the affected and the contralateral sides is an important outcome, because it is independent of age and sex.[26-29]

For the quantification of reproducibility, we calculated the absolute amount of measurement error (agreement) and the ability of examiners to differentiate between different subjects (reliability).[30-32]

Agreement. The mean difference between the 2 examiners and the SD of this difference was calculated. When the mean difference deviates substantially from 0°, it indicates a systematic difference between the examiners. Although there are no clear criteria for the acceptable degree of interobserver agreement, we considered differences exceeding 10° to be low agreement. We then calculated the percentage of differences between the 2 examiners equal to or lower than 10°.

The magnitude of the SD expresses the extent to which the examiners are able to achieve the same value.[30,33] Subsequently, the 95% limits of agreement were calculated, defined as the mean difference between examiners ±1.96 × SDs of this mean difference. Only differences between examiners that exceed the limits of agreement can be interpreted as "real" differences above measurement error.[34]

Based on analysis of variance (ANOVA), the standard error of measurement (SEM) was calculated as an additional measure of agreement. The SEM was defined as $\sqrt{(\sigma^2_{pt}+\sigma^2_{residual})}$ with σ^2_{pt} referring to the variance due to systematic differences between the examiners. We included σ^2_{pt} because we were interested in absolute agreement between the examiners.

Reliability. The intraclass correlation coefficient (ICC) was derived from a 2-way random-effects ANOVA (model 2 according to the guidelines specified by Fleiss[35]). An ICC of at least .70 is considered satisfactory for group comparisons.[36]

Influence of clinical characteristics on the level of agreement.

To test whether the level of interobserver agreement was influenced by clinical characteristics, the limits of agreement were calculated for different predefined subgroups: bilateral involvement (yes, no) and duration of complaints (≤6mo, [table 1] >6mo). For continuous variables, such as pain (patient's score for the past 24h, mean VAS score of the 2 examiners) and disability (SDQ score), the 25th and 75th percentiles were used to form 3 subgroups: low, moderate, and high severity of pain or disability. Because sound comparison of ICC values is only possible in subgroups with a similar level of heterogeneity,[32] subgroup analysis was performed only for the limits of agreement.

RESULTS

Table 1 shows the characteristics of the study population. The duration of shoulder complaints varied considerably. Sleep disturbances and pain at rest were reported frequently. Furthermore, 38% of the patients were unable to perform their normal daily activities.

Reproducibility of the Visual Estimation of ROM

The results of the interobserver agreement and reliability with regard to the visual estimation of ROM are presented in table 2. The lowest level of agreement was found for the visual estimation of active and passive elevation (-43.4 to 39.8 and -46.7 to 41.5, respectively, for the difference between the affected and contralateral sides). This was also reflected in a higher SEM. For the passive external rotation, relatively large systematic differences were found for the affected and contralateral sides, but no significant systematic differences were found for the difference between the affected and contralateral sides.

For all assessments, higher level of interobserver agreement was found for the contralateral side, expressed by smaller limits of agreement. Furthermore, the percentage of differences within 10° was higher for most movements of the contralateral side.

The ICCs were below .70 for the visual estimation of ROM during horizontal adduction. For the other movements, ICCs of at least .70 were found for the difference between the affected and contralateral sides.

Influence of Clinical Characteristics on the Level of Agreement

In table 3, the influence of pain (mean VAS score of the examiners) and disability (SDQ total score) on the limits of agreement are presented for the assessment of the difference between the affected and contralateral sides. Subgroup analysis showed that for high pain severity (mean pain score, >72) and a high level of disability (SDQ score, >87), the limits of agreement were larger than in the subgroups with moderate and low severity of complaints, especially in active and passive elevation. Similar results were found for the assessment of pain by the patient (data not shown). We found no effect on the limits of agreement for duration of complaints or bilateral involvement.

DISCUSSION

In this study, the lowest level of agreement was the visual estimation of active and passive elevation, movements for which the level of agreement was most clearly associated with the severity of pain and disability. The ability to differentiate between subjects was acceptable for all movements for the difference between the affected and contralateral sides, except for passive horizontal adduction.

We found relatively large systematic differences between the 2 examiners for the scapulohumeral abduction and external rotation. Croft et al[21] found that differences in the physical examination technique can account for systematic differences between the examiners for the external rotation. We found relatively large systematic differences for external rotation in our results of the affected and the contralateral sides, compared with the assessment of the difference between the

affected and the contralateral sides. Therefore, these systematic differences might also be explained by differences in the physical examination technique, such as the amount of force applied or the degree of fixation of the trunk. There is no apparent explanation for the systematic differences found for scapulohumeral abduction.

In agreement with the results of Boström et al,[13] we found that the lowest level of interobserver agreement was for elevation. How can this low level of interobserver agreement for elevation be explained? During elevation, several anatomic structures are susceptible to compression. Compared with the other movements, we found that elevation was more frequently associated with pain and, furthermore, patients considered movements above shoulder level to be the main aggravating factor (data not shown). Therefore, it seems reasonable that as a consequence of the pain experienced during elevation and fear of overloading the shoulder, the results of the elevation might vary more than the results of other movements.

Given the large limits of agreement for elevation, one could question whether an effect of repeated examination might also explain the results. Several patients reported that, because of the physical examination performed by the first examiner, pain intensity during the second examination was increased. In another analysis, an effect of repeated examination was found for the elevation but not for the other movements; this would indicate a tendency toward more restriction during the second examination, especially in patients with more severe pain (data not shown).

In this study, the level of agreement was not influenced by bilateral involvement or chronic complaints but was dependent on the severity of pain and disability. The limits of agreement were relatively large for patients with severe pain or disability. For example, if the active elevation of a patient with a low level of pain is measured by 2 examiners, there is a 95% probability that the difference between the 2 examiners will be between -13.9° and 14.0° as a sole effect of measurement error. For a patient with a high level of pain, this range is -74.6° and 60.9° (see table 3). Only differences that exceed the limits of agreement can be interpreted as “real” differences above measurement error.

[TABLE 2]

As a consequence of the influence of pain and disability, it is difficult to compare the results on reproducibility of different studies with varying populations. It has been suggested that quantification of the ROM using a goniometer, inclinometer, or some other device would be more reliable than visual estimation of the ROM.[22] However, several studies[37-42] using different devices (eg, visual estimation, inclinometer, FASTRAK system) recently reported levels of agreement and reliability that were similar to our findings. A comparative study would be needed to draw conclusions about the value of visual estimation of the ROM compared with the use of a measurement device.

Another implication of our results is that it seems attractive to test ROM until pain is present to avoid overloading the affected shoulder. However, it has been shown that this strategy is less reproducible than testing to the point of maximum elevation.[2] This finding was based on a small study population and more research is needed to determine whether this is beneficial in improving the reproducibility. Based on our findings— that is, that repeated physical examination did increase the pain experienced by patients—and the fact that high pain severity increases the differences between examiners, we recommend that the physical examination be restricted to only a few movements in patients with a high pain severity.

[TABLE 3]

Our study shows that by using visual estimation, one can adequately discriminate between groups of patients with different ROMs for most movements, except for passive horizontal adduction. However, for discrimination between individual patients in clinical practice, the reliability of all measurements was somewhat low.

To be useful for outcome assessment in clinical practice or research, an instrument should have high responsiveness, which is strongly determined by the level of agreement.[43] The limits of agreement should be smaller than the minimal clinically important difference that one wants to detect. This judgment should be made separately for each application of the method. Given the large limits of

agreement, the value of active and passive elevation as an outcome measure can be questioned when different examiners are involved in the assessment. However, often only 1 examiner per patient will be involved in the assessment of outcome. Unfortunately, practical reasons made it impossible to investigate the level of intraobserver reproducibility in this study. Generally, the intraobserver reproducibility will be higher than the interobserver reproducibility.[12,13,15] Future studies should investigate whether patient characteristics also have a considerable effect on the level of intraobserver agreement. Furthermore, it would be useful to investigate the responsiveness of visual estimation of the shoulder ROM.

CONCLUSIONS

Interobserver agreement was low in the assessment of active and passive elevation, movements for which the magnitude of the limits of agreement clearly increased with higher severity of pain and disability. Thus the value of visual estimation as an outcome measure can be questioned when different examiners are involved. Except for horizontal adduction, visual estimation of the ROM seems to be a suitable method for distinguishing differences between the affected and contralateral ROM of subjects. In future studies, the value of visual estimation should be compared with other measurement methods, and the responsiveness should be assessed.

TABLES

Table 1: Characteristics of the Study Population (N=201)

Female (%)	66
Mean age ± SD (y)	48±12
Dominant shoulder affected (%)	54
Bilateral shoulder complaints (%)	27
Previous episode(s) of shoulder complaints (%)	40
Duration of current episode (%)	
<3mo	27
3–6mo	16
6–12mo	22
>12mo	35
Pain at rest (%)	67
Sleep disturbances (%)	
Unable to lie on the involved shoulder	51
Waking up; cannot fall asleep	69
Unable to perform normal daily activities (%)	38
Mean SDQ score ± SD*	68±23
Mean pain score ± SD†	50±26
Patients recruited by (%)	
General practitioners	37
Physicians in orthopedic practice	17
Rheumatologists in secondary care rheumatology clinic	46

*SDQ score (0–100); higher score indicates a higher level of disability.
†Severity of pain assessed independently by both examiners on a VAS (0–100); higher score indicates more severe pain. The mean scores of both examiners are presented.

Table 2: Reproducibility of the Visual Estimation of Shoulder ROM

Examined Movements	Examiner A (deg) Mean ± SD	Examiner B (deg) Mean ± SD	Agreement: Examiner A – B (deg) Mean ± SD	Limits of Agreement*	Agreement (%) Within 10°	Agreement SE (95% CI)	Reliability ICC (95% CI)
Active elevation							
Affected side	144.5±41.9	139.4±42.9	4.7±20.1	−34.7 to 44.1	68	14.6	.88 (.84–.91)
Contralateral side	173.4±11.7	170.4±14.7	2.9±8.8	−14.3 to 20.1	89	6.6	.76 (.67–.82)
Contralateral – affected	28.9±38.9	31.0±39.6	−1.8±21.2	−43.4 to 39.8	71	15.0	.85 (.81–.89)
Passive elevation							
Affected side	145.9±44.5	141.8±44.3	4.1±22.7	−40.4 to 48.6	63	16.2	.87 (.83–.90)
Contralateral side	176.0±15.1	174.6±16.2	1.3±11.4	−21.0 to 23.6	90	8.1	.73 (.66–.79)
Contralateral – affected	29.9±41.1	32.5±40.0	−2.6±22.5	−46.7 to 41.5	59	16.0	.85 (.80–.88)
Passive scapulohumeral abduction‡							
Affected side	69.6±20.0	79.8±18.3	−9.2±13.4	−35.5 to 17.1	60	11.5	.67 (.35–.81)
Contralateral side	84.1±11.8	89.5±5.8	−4.9±11.6	−27.6 to 17.8	88	8.9	.15 (.02–.29)
Contralateral – affected	14.5±19.6	9.4±17.3	4.5±13.8	−22.5 to 31.5	77	10.3	.70 (.59–.78)
Passive external rotation							
Affected side	64.6±22.4	53.0±19.6	11.2±12.0†	−12.3 to 34.7	49	11.6	.73 (.22–.88)
Contralateral side	78.9±9.0	64.3±12.8	14.6±9.5†	−4.0 to 33.2	43	12.3	.34 (.00–.65)
Contralateral – affected	14.4±20.1	11.3±17.8	3.4±12.4	−20.9 to 27.7	73	9.1	.77 (.69–.83)
Passive horizontal adduction‡							
Affected side	130.7±16.1	129.4±12.6	1.5±16.1	−30.1 to 33.1	75	11.4	.36 (.22–.48)
Contralateral side	137.0±12.7	135.9±7.5	0.5±13.5	−25.5 to 27.5	86	9.5	.18 (.04–.32)
Contralateral – affected	6.5±10.6	6.7±9.9	−0.5±9.8	−19.8 to 18.7	87	7.0	.49 (.37–.60)

Abbreviation: CI, confidence interval.
*Calculated as the mean difference between the examiners ±1.96 × SDs of this mean difference.
†Greater than 10° difference.
‡For the passive scapulohumeral abduction, no data were available for 20 (10%) patients because palpation of the scapular movement was not possible due to adipositas or insufficient relaxation of the patients. For the horizontal passive adduction, no data were available for 31 patients (15%) because of high pain intensity or the fact that 90° anteflexion was impossible.

Table 3: Influence of Pain and Disability on the Limits of Agreement

Examined Movements*	Pain Severity† Low (VAS <30)	Pain Severity† Moderate (30≤VAS≤72)	Pain Severity† High (VAS >72)	Severity of Disability‡ Low (SDQ <54)	Severity of Disability‡ Moderate (54≤SDQ≤87)	Severity of Disability‡ High (SDQ >87)
Active elevation	–13.9 to 14.0	–33.4 to 32.6	–74.6 to 60.9	–19.4 to 21.2	–33.0 to 27.8	–71.5 to 65.7
Passive elevation	–20.3 to 17.2	–46.4 to 38.8	–64.0 to 59.3	–34.6 to 21.5	–45.0 to 42.3	–58.6 to 57.1
Passive scapulohumeral abduction	–20.6 to 32.1	–23.2 to 30.1	–25.2 to 34.9	–17.7 to 23.5	–25.3 to 36.6	–21.3 to 28.4
Passive external rotation	–17.1 to 18.4	–20.6 to 29.5	–23.9 to 32.0	–21.3 to 25.3	–20.0 to 26.7	–21.4 to 32.2
Passive horizontal adduction	–7.5 to 9.8	–21.5 to 20.2	–29.1 to 24.3	–18.2 to 16.5	–20.2 to 19.6	–21.5 to 20.0

*Only the results of the difference between contralateral side and affected sides are presented.
†For mean VAS pain score of both examiners, the values of the 25th and 75th percentiles were used to form subgroups.
‡For SDQ score, the values of the 25th and 75th percentiles were used to form subgroups.

REFERENCES

1. Pope DP, Croft PR, Pritschard CM, Macfarlane GJ, Silman AJ. The frequency of restricted range of movement in individuals with self-reported shoulder pain: results from a population-based survey. Br J Rheumatol 1996;35:1137-41
2. Clarke GR, Willis LA, Fish WW, Nichols PJ. Preliminary studies in measuring range of motion in normal and painful stiff shoulders. Rheumatol Rehabil 1975;14:39-46
3. Hedtmann A, Fett H. [So-called humero-scapular periarthropathy— classification and analysis based on 1,266 cases] [German]. Z Orthop Ihre Grenzgeb 1989;127:643-9
4. Bjelle A. Epidemiology of shoulder problems. Baillière's Clin Rheumatol 1989;3:437-51
5. Triffitt DP. The relationship between motion of the shoulder and the stated ability to perform activities of daily living. J Bone Joint Surg Am 1998;80:41-6
6. Badley EM, Wagstaff S, Wood PH. Measures of functional ability (disability) in arthritis in relation to impairment of range of motion. Ann Rheum Dis 1984;43:563-9
7. Vecchio PC, Kavanagh RT, Hazleman BL, King RH. Community survey of shoulder disorders in the elderly to assess the natural history and effects of treatment. Ann Rheum Dis 1995;54:152-4
8. Cyriax J. Textbook of orthopaedic medicine. 7th ed. London: Bailliere Tindall; 1981. p 190-239
9. Green S, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. BMJ 1998;316:354-60

10. Green S, Buchbinder R, Forbes A, Bellamy N. A standardized protocol for measurement of range of movement of the shoulder using the pluri-V inclinometer and assessment of its intrarater and interrater reliability. Arthritis Care Res 1998;11:43-52
11. Heemskerk MA, van Aarst M, van der Windt DA. De reproduceerbaarheid van het meten van de passieve beweeglijkheid van de schouder met de EDI-320 digitale hoekmeter. Ned T Fysiother 1997;107:146-9
12. Riddle DL, Rothstein JM, Lamb RL. Goniometric reliability in a clinical setting; shoulder measurements. Phys Ther 1987;67:668- 73
13. 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
14. van den Ende CH, Rozing PM, Dijkmans BA, Verhoef JA, Voogtvan der Harst EM, Hazes JM. Assessment of shoulder function in rheumatoid arthritis. J Rheumatol 1996;23:2043-8
15. Johnson GR, Fyfe NC, Heward M. Ranges of movement at the shoulder complex using an electromagnetic movement sensor. Ann Rheum Dis 1991;50:824-7
16. Laupattarakasem W, Sirichativapee W, Kowsuwon W, Sribunditkul S, Suibnugarn C. Axial rotation gravity goniometer; a simple design of instrument and a controlled reliability study. Clin Orthop 1990;Feb(251):271-4
17. Youdas JW, Carey JR, Garrett TR, Suman VJ. Reliability of goniometric measurements of active arm elevation in the scapular plane obtained in a clinical setting. Arch Phys Med Rehabil 1994;75:1137-44
18. Doody SG, Waterland JC, Freedman L. Scapulo-humeral goniometer. Arch Phys Med Rehabil 1970;51:711-3
19. Sabari JS, Maltzev I, Lubarsky D, Liszkay E, Homel P. Goniometric assessment of shoulder range of motion: comparison of testing in supine and sitting positions. Arch Phys Med Rehabil 1998;79:647-51
20. Williams JG, Callaghan M. Comparison of visual estimation and goniometry in determination of a shoulder joint angle. Physiotherapy 1990;76:655-7
21. Croft P, Pope D, Boswell R, Rigby A, Silman A. Observer variability in measuring elevation and external rotation of the shoulder. Br J Rheumatol 1994;33:942-6
22. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther 1987;67:1867-72
23. de Winter AF, Jans MP, Scholten RJ, Devillé W, van Schaardenburg D, Bouter LM. Diagnosis of shoulder pain: inter-observer agreement and determinants of diagnostic disagreement. Ann Rheum Dis 1999;58:272-7
24. Bakker JF, de Jongh L, Jonguiere M, Mens J, Oosterhuis WW, Poppelaars A. Standaard schouderklachten. Huisarts Wetenschap 1990;33:196-202
25. van der Windt DA, van der Heijden GJ, de Winter AF, Koes BW, Devillé W, Bouter LM. The responsiveness of the Shoulder Disability Questionnaire. Ann Rheum Dis 1998;57:82-7
26. van der Heijden GJ, Leffers P, Bouter LM. Shoulder disability questionnaire: design and responsiveness of a functional status measure. J Clin Epidemiol 2000;53:29-38
27. van Schaardenburg D, Van den Brande KJ, Ligthart GJ, Breedveld FC, Hazes JM. Musculoskeletal disorders in persons aged 85 and over: a community survey. Ann Rheum Dis 1994;53:807-11
28. Bassey EJ, Morgan K, Dallasso HM, Ebrahim SB. Flexibility of the shoulder joint measured as range of abduction in a large representative sample of men and women over 65 years of age. Eur J Physiol 1989;58:353-60
29. Allander O, Bjornsson OJ, Olafsson O, Sigfusson N, Thorsteinsson J. Normal range of movements in shoulder, hip, wrist and thumb with a special reference to side; a comparison between populations. Int J Epidemiol 1974;3:253-61
30. de Vet HC. Observer reliability and agreement. In: Armitage P, Colton T, editors. Encyclopedia of biostatistics. Vol 4. New York: John Wiley & Sons; 1998. p 3123-8
31. Guyatt GH, Kirshner B, Jaeschke R. Measuring health status: what are the necessary measurement properties? J Clin Epidemiol 1992;45:1341-5
32. Stratford P. Reliability: consistency or differentiating between subjects? Phys Ther 1989;69:299-300
33. Tinsley HE, Weiss DJ. Interrater reliability and agreement of subjective judgements. J Couns Psychol 1975;22:358-76
34. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1:307-10

35. Fleiss JF. The design and analysis of clinical experiments. New York: John Wiley & Sons; 1986. p 1-33
36. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996;18:979-92
37. Green S, Buchbinder R, Forbes A, Bellamy N. A standardized protocol for measurement of range of movement of the shoulder using the Plurimeter-V inclinometer and assessment of its intrarater and interrater reliability. Arthritis Car Res 1998;11: 43-52
38. Triffitt PD, Wildin C, Hajioff D. The reproducibility of measurement of shoulder movement. Acta Orthop Scand 1999;70:322-4
39. Jordan K, Dziedzic K, Jones PW, Ong BN, Dawes PT. The reliability of the three-dimensional FASTRAK measurement system in measuring cervical spine and shoulder range of motion in healthy subjects. Rheumatology 2000;39:382-8
40. Hayes K, Walton JR, Szomor ZR, Murrell GA. Reliability of five methods for assessing shoulder range of motion. Aust J Physiother 2001;47:289-94
41. Hoving JL, Buchbinder R, Green S, et al. How reliably do rheumatologists measure shoulder movement? Ann Rheum Dis 2002; 61:612-6
42. de Winter AF, Heemskerk MA, Terwee CB, et al. Inter-observer reproducibility of range of motion in patients with shoulder pain using a digital inclinometer. BMC Musculoskelet Disord 2004; 5:18
43. de Vet HC, Bouter LM, Bezemer PD, Beurskens AJ. Reproducibility and responsiveness of evaluative outcome measures. Theoretical considerations illustrated by an empirical example. Int J Technol Assess Health Care 2001;17:479-87.