How reliably do rheumatologists measure shoulder movement?

J L Hoving, R Buchbinder, S Green, A Forbes, N Bellamy, C Brand, R Buchanan, S Hall, M Patrick, P Ryan, A Stockman


Objective: To assess the intrarater and interrater reliability among rheumatologists of a standardised protocol for measurement of shoulder movements using a gravity inclinometer.

Methods: After instruction, six rheumatologists independently assessed eight movements of the shoulder, including total and glenohumeral flexion, total and glenohumeral abduction, external rotation in neutral and in abduction, internal rotation in abduction and hand behind back, in random order in six patients with shoulder pain and stiffness according to a 6x6 Latin square design using a standardised protocol. These assessments were then repeated. Analysis of variance was used to partition total variability into components of variance in order to calculate intraclass correlation coefficients (ICCs).

Results: The intrarater and interrater reliability of different shoulder movements varied widely. The movement of hand behind back and total shoulder flexion yielded the highest ICC scores for both intrarater reliability (0.91 and 0.83, respectively) and interrater reliability (0.80 and 0.72, respectively). Low ICC scores were found for the movements of glenohumeral abduction, external rotation in abduction, and internal rotation in abduction (intrarater ICCs 0.35, 0.43, and 0.32, respectively), and external rotation in neutral, external rotation in abduction, and internal rotation in abduction (intrarater ICCs 0.29, 0.11, and 0.06, respectively).

Conclusions: The measurement of shoulder movements using a standardised protocol by rheumatologists produced variable intrarater and interrater reliability. Reasonable reliability was obtained only for the movement of hand behind back and total shoulder flexion.

Shoulder pain is common in the general population, its point prevalence averages between 7% and 51% and it is known to increase with age. Restricted range of motion and shoulder pain can interfere with activities in daily life and is associated with work absenteeism and use of medical services. Many patients receive some evaluation by a family doctor, rheumatologist, orthopaedic specialist, or physical therapist.

A physical examination is often used for both diagnosis and evaluation of treatment success in patients with shoulder pain. One aspect of physical assessment of the shoulder is the evaluation of range of motion. No “gold standard” for the measurement of shoulder range of motion is yet available. Clinical trials that have assessed the efficacy of interventions for shoulder pain have commonly used range of motion as a measurement tool. To be of value in clinical trials or routine care its reliability (that is, the repeated administration of an instrument to a stable population yielding the same results) should be established.

Multiple methods for estimating shoulder range of motion have been used in the past, including visual estimation, the two armed goniometer, or a gravity referenced goniometer. In many of these studies the methods are poorly described and most looked at passive range of motion only, which may reflect a less functional outcome. The results may not be applicable to symptomatic patients with varying degrees of pain and stiffness. Only a few studies have included subjects with shoulder pain.

We previously developed a standardised protocol, which we assessed in a similar group, for the measurement of active range of motion in clinical trials for shoulder pain. The intrarater and interrater reliability of our standardised protocol was found to be acceptable when physiotherapists performed the range of motion assessments. We concluded that it would be appropriate for use both in research and clinical practice by physiotherapists. The gravity inclinometer chosen is fast and easy to use but has not been evaluated in observers with a different professional background.

To determine whether our protocol may also be of value when rheumatologists perform the measurements, this study aimed at examining the intrarater reliability (that is reliability when performed by multiple raters) and interrater reliability (that is reliability when performed again and again by the same rater or examiner) of the standardised protocol when performed by six rheumatologists. Although this study replicates a previous study, which looked at different subjects with similar complaints, differences in professional training and daily use of manual examination techniques between physiotherapists and rheumatologists might influence the reliability of the measurements. For this reason, we compared the results with those obtained by six physiotherapists.

METHODS

The standardised protocol for the measurement of shoulder movement with the Plurimeter-V gravity inclinometer (Dr Rippstein, Zurich, Switzerland) was described in detail previously. To minimise potential sources of variation, for each movement, we specified the position of the patient, positioning of the instrument on the patient, stabilisation of the joint to be measured, and determination of the end point of movement. The gravity referenced inclinometer was used to measure in total eight movements: total shoulder flexion; glenohumeral flexion; total and glenohumeral abduction; internal rotation; and external rotation in both neutral and abduction. Internal rotation was performed in abduction only. Hand behind back, assessed by palpation to the nearest spinal process, was also included.
Six patients with varying degrees of pain and stiffness in the shoulder, and varying diseases, were recruited from a private rheumatology practice. Subjects were excluded if repeated testing was likely to aggravate the condition.

Six rheumatologists undertook a one hour training session to familiarize themselves with the process and standardize the measurement methods between the examiners. All were given a detailed manual. All had experience in measuring shoulder range of motion, but none had used an inclinometer. The examiners in the study were instructed to demonstrate each movement to the patient and ask the patient to move the arm until the first sensation of pain so that each examiner agreed on the definition of the first onset of pain. Care was taken in aligning the inclinometer on the patient’s arm, ensuring consistent positioning of adjacent joints. The participant baseline data collected included age, sex, duration of the shoulder pain, a detailed manual. All had experience in measuring shoulder range of motion, but none had used an inclinometer. The examiners in the study were instructed to demonstrate each movement to the patient and ask the patient to move the arm until the first sensation of pain so that each examiner agreed on the definition of the first onset of pain. Care was taken in aligning the inclinometer on the patient’s arm, ensuring consistent positioning of adjacent joints. The participant baseline data collected included age, sex, duration of the shoulder pain, and a history of trauma or shoulder surgery. In addition, the examiners were unaware of each other’s scores and their own scores of the previous round. Between rounds 1 and 2 there was a one hour break for both patients and examiners. The order of the examiners and the order of the eight movements were randomly assigned for each round separately. In addition, the examiners were in separate rooms, and had no contact with the other examiners during the measurement procedures. Thus the examiners were unaware of each other’s scores and their own scores of the previous round.

### Statistical analyses

Analysis of variance was used to partition the total variability into components of variation due to patients, examiners, round, two-way interactions of individual components, and residual error (using the SAS V ARCOMP procedure). All sources were considered as random effects.

Results are presented as the total variability and the percentage of the total variability attributable to each variance component. Intrarater and interrater reliability were calculated from the variance (var) components through appropriate intraclass correlation coefficients (ICCs) according to the following formulae:

\[ \text{Intrarater reliability} = \frac{\text{var(patient)} + \text{var(observer)} + \text{var(patient}\times\text{observer})}{\text{var(total)}} \]

### Table 1: Summary statistics for each movement measured in six patients in rounds 1 and 2 by the six rheumatologists

<table>
<thead>
<tr>
<th>Round</th>
<th>Patient 1 Mean (SD)</th>
<th>Patient 2 Mean (SD)</th>
<th>Patient 3 Mean (SD)</th>
<th>Patient 4 Mean (SD)</th>
<th>Patient 5 Mean (SD)</th>
<th>Patient 6 Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>102.7 (11.2)</td>
<td>136.5 (12.1)</td>
<td>131.7 (21.7)</td>
<td>66.7 (22.3)</td>
<td>131.3 (5.0)</td>
<td>133.0 (12.4)</td>
</tr>
<tr>
<td>2</td>
<td>89.2 (8.5)</td>
<td>134.7 (12.2)</td>
<td>136.0 (13.7)</td>
<td>80.5 (29.4)</td>
<td>111.5 (36.6)</td>
<td>91.0 (14.8)</td>
</tr>
<tr>
<td>1</td>
<td>71.0 (7.0)</td>
<td>100.3 (13.1)</td>
<td>97.3 (18.4)</td>
<td>57.7 (12.2)</td>
<td>77.5 (16.2)</td>
<td>85.0 (29.2)</td>
</tr>
<tr>
<td>2</td>
<td>62.3 (16.4)</td>
<td>95.6 (21.0)</td>
<td>123.3 (16.0)</td>
<td>45.5 (13.1)</td>
<td>73.3 (25.7)</td>
<td>85.0 (31.6)</td>
</tr>
<tr>
<td>1</td>
<td>68.2 (17.8)</td>
<td>118.2 (29.7)</td>
<td>99.7 (15.6)</td>
<td>77.5 (16.2)</td>
<td>85.0 (29.2)</td>
<td>85.0 (31.6)</td>
</tr>
<tr>
<td>2</td>
<td>74.8 (14.8)</td>
<td>127.0 (18.3)</td>
<td>140.0 (29.7)</td>
<td>73.3 (25.7)</td>
<td>85.0 (29.2)</td>
<td>85.0 (31.6)</td>
</tr>
<tr>
<td>1</td>
<td>46.7 (15.4)</td>
<td>52.8 (16.0)</td>
<td>74.6 (27.0)</td>
<td>37.7 (11.7)</td>
<td>57.3 (10.0)</td>
<td>57.3 (10.0)</td>
</tr>
<tr>
<td>2</td>
<td>67.0 (6.4)</td>
<td>82.3 (21.1)</td>
<td>107.7 (28.2)</td>
<td>49.7 (25.3)</td>
<td>62.7 (15.3)</td>
<td>62.7 (15.3)</td>
</tr>
<tr>
<td>1</td>
<td>40.0 (16.6)</td>
<td>31.8 (9.8)</td>
<td>46.6 (23.1)</td>
<td>46.2 (12.4)</td>
<td>39.8 (24.1)</td>
<td>39.8 (24.1)</td>
</tr>
<tr>
<td>2</td>
<td>38.2 (20.4)</td>
<td>35.0 (13.7)</td>
<td>42.2 (21.3)</td>
<td>52.3 (16.2)</td>
<td>35.3 (18.1)</td>
<td>35.3 (18.1)</td>
</tr>
<tr>
<td>1</td>
<td>35.7 (14.7)</td>
<td>32.7 (19.8)</td>
<td>34.3 (21.1)</td>
<td>62.5 (41.5)</td>
<td>35.2 (25.9)</td>
<td>35.2 (25.9)</td>
</tr>
<tr>
<td>2</td>
<td>41.0 (17.7)</td>
<td>42.3 (8.1)</td>
<td>46.8 (26.7)</td>
<td>49.5 (26.8)</td>
<td>21.3 (10.6)</td>
<td>21.3 (10.6)</td>
</tr>
<tr>
<td>1</td>
<td>54.3 (16.5)</td>
<td>64.8 (24.9)</td>
<td>70.0 (31.9)</td>
<td>78.5 (21.4)</td>
<td>65.3 (22.1)</td>
<td>65.3 (22.1)</td>
</tr>
<tr>
<td>2</td>
<td>54.7 (22.7)</td>
<td>80.7 (23.0)</td>
<td>60.5 (25.7)</td>
<td>57.8 (26.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11.0 (8.0–15.0)</td>
<td>9.5 (3.0–12.0)</td>
<td>9.0 (4.0–11.0)</td>
<td>7.0 (6.0–10.0)</td>
<td>19.0 (17.0–19.0)</td>
<td>19.0 (17.0–19.0)</td>
</tr>
<tr>
<td>2</td>
<td>11.0 (10.0–13.0)</td>
<td>8.0 (4.0–9.0)</td>
<td>8.0 (4.0–11.0)</td>
<td>8.0 (6.0–10.0)</td>
<td>19.0 (17.0–19.0)</td>
<td>19.0 (17.0–19.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Movement</th>
<th>Intrarater ICC (95% CI)*</th>
<th>Interrater ICC (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSF</td>
<td>0.83 (0.52 to 0.94)</td>
<td>0.72 (0.33 to 0.93)</td>
</tr>
<tr>
<td>GHF</td>
<td>0.62 (0.18 to 0.83)</td>
<td>0.65 (0.32 to 0.86)</td>
</tr>
<tr>
<td>TSA</td>
<td>0.56 (0.21 to 0.73)</td>
<td>0.49 (0.03 to 0.86)</td>
</tr>
<tr>
<td>GHA</td>
<td>0.35 (0.04 to 0.65)</td>
<td>0.51 (0.22 to 0.77)</td>
</tr>
<tr>
<td>EBA</td>
<td>0.75 (0.42 to 0.89)</td>
<td>0.29 (0.02 to 0.68)</td>
</tr>
<tr>
<td>ERA</td>
<td>0.43 (0.12 to 0.60)</td>
<td>0.11 (0.02 to 0.38)</td>
</tr>
<tr>
<td>EBB</td>
<td>0.32 (0.02 to 0.67)</td>
<td>0.06 (0.01 to 0.45)</td>
</tr>
<tr>
<td>HBB</td>
<td>0.91 (0.41 to 0.97)</td>
<td>0.80 (0.19 to 0.95)</td>
</tr>
</tbody>
</table>

**TSF**, total shoulder flexion; **GHF**, glenohumeral flexion; **TSA**, total shoulder abduction; **GHA**, glenohumeral abduction; **ERN**, external rotation in neutral; **ERA**, external rotation in abduction; **IRA**, internal rotation in abduction; **HBB**, hand behind back; **SD**, standard deviation. For **HBB** median and range are presented instead of mean and SD.
Interrater reliability = \( \frac{\text{var}(\text{patient}) + \text{var}(\text{round}) + \text{var}(\text{patient} \times \text{round})}{\text{var}(\text{total})} \)

The closer the ICC is to 1, the better the reliability. Statistical uncertainty in the computed ICCs is expressed through approximate 95% confidence intervals obtained from 5000 bootstrap resamples of both subjects and rheumatologists.

Interexaminer reliabilities were also calculated for each round separately:

\[
\text{Interrater reliability} = \frac{\text{var}(\text{patient})}{\text{var}(\text{total})}
\]

RESULTS
Six patients (two male, four female), with a mean age of 72 years (range 54–82), took part in the study. The right shoulder was assessed in four patients and the left shoulder in two. The mean duration of symptoms was 16 months (range 1–48). Two patients had a history of shoulder trauma and one patient had previous shoulder surgery. The severity rating by each patient ranged from 2 (mild severity) to 3 (moderate severity) on the ordinal five point scale.

Table 1 shows the means and standard deviations for each of the six patients in rounds 1 and 2, combining the scores of the six rheumatologists for each movement. The range of motion of patients varied considerably. For example, the mean total shoulder flexion in round 1 varied between 66.7° (SD 22.3) and 136.5° (12.1). The differences between rounds 1 and 2 were smaller.

Table 2 shows the intrarater and interrater ICCs calculated for each of the eight shoulder movements together with their approximate 95% confidence intervals. The movement of hand behind back and total shoulder flexion gave the highest ICC scores for both intrarater reliability (0.91 and 0.83, respectively) and interrater reliability (0.80 and 0.72, respectively). Low ICC scores were found for the movements of glenohumeral abduction, external rotation in abduction, and internal rotation in abduction (intrarater ICCs 0.35, 0.43, and 0.32, respectively); and external rotation in neutral, external rotation in abduction, and internal rotation in abduction (interrater ICCs 0.29, 0.11 and 0.06, respectively).

Table 3 gives the individual components of variation as a percentage of the total variance. The proportion of the variance between examiners for the movements of external rotation in neutral, external rotation in abduction, and internal rotation in abduction was considerably greater than that of the explained variance of the patients, thus yielding very low interrater ICCs (see respective formula) for these movements.

In a previous study, we evaluated the reliability of shoulder measurement between physiotherapists using the same design and a similar patient group with shoulder pain. Figures 1 and 2 display graphically the ICCs of the rheumatologists compared with those of the physiotherapists. Overall, the physiotherapists achieved higher intrarater and interrater reliability ratings than rheumatologists, especially for the movements of external rotation and internal rotation.

---

| Table 3 Percentage of total variance attributable to patient, observer, round, two-way interactions, residual error and total variance for shoulder movements |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | Patient (%)     | Observer (%)    | Round (%)       | Patient by round (%) | Round by observer (%) | Patient by observer (%) | Residual error (%) | Total variance |
| TSF             | 71              | 2               | 0               | 1                | 0                | 10              | 16              | 1007.6         |
| GHF             | 52              | 4               | 0               | 13               | 4                | 7               | 21              | 867.6          |
| TSA             | 41              | 14              | 3               | 5                | 0                | 0               | 37              | 1138.1         |
| GHA             | 30              | 5               | 17              | 5                | 0                | 0               | 44              | 743.7          |
| ERN             | 29              | 38              | 0               | 0                | 4                | 8               | 21              | 446.8          |
| ERA             | 8               | 36              | 0               | 3                | 0                | 0               | 54              | 667.4          |
| IRA             | 6               | 26              | 0               | 0                | 0                | 0               | 68              | 640.2          |
| HBB             | 79              | 1               | 0               | 0                | 0                | 10              | 9               | 21.7           |

TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back.

---

**Figure 1** Intrarater reliability of rheumatologists and physiotherapists for each movement. TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back.

**Figure 2** Interrater reliability of rheumatologists and physiotherapists for each movement. TSF, total shoulder flexion; GHF, glenohumeral flexion; TSA, total shoulder abduction; GHA, glenohumeral abduction; ERN, external rotation in neutral; ERA, external rotation in abduction; IRA, internal rotation in abduction; HBB, hand behind back.

---

www.annrheumdis.com
DISCUSSION
In this study we found that using a standardised protocol for the measurement of shoulder movements produced variable intrarater and interrater reliability. Reasonable reliability (ICC > 0.7) was obtained for the movement of hand behind back and total shoulder flexion only. The other movements in our protocol produced fair to poor reliability, especially for those movements demanding more complex stabilisation and handling skills, such as external rotation and internal rotation. There was a learning effect as demonstrated by the increase in interrater reliability in round 2, though we did not formally test for statistical significance of this apparent increase in reliability. However, we think that the reliability can be improved with practice.

As far as we know, no previous published studies have investigated the differences in reliability between different groups of raters for measuring shoulder movements. Physiotherapists and rheumatologists may differ in their use of manual techniques, especially those movements requiring fixation, which are performed more often by physiotherapists. There may also be other determinants of the differences other than the professional background of the raters, such as experience in examining shoulders, different verbal and non-verbal cues to patients, and level of confidence, which may also have a role. The differences in reliability might also depend on the selection of the individual raters. For example, the physiotherapists selected in our previous study might have been a more homogeneous group. In addition, we cannot exclude the possibility that the observed differences between the ICCs of the physiotherapists and rheumatologists were due to real differences between the patients. Although both studies recruited patients with stiff painful shoulders with similar duration of symptoms and pain scores, some minor differences were noted, including the fact that patients in this study had more homogeneous pain scores. Unfortunately, the small number of patients in our study does not allow for more analyses to examine the sources of disagreement between the rheumatologists and physiotherapists.

Other studies have reported interrater ICC values for patients with shoulder disorders ranging between 0.26 and 0.95.7-15 The reliability scores in our study are consistent with the overall conclusion of previous studies that reliability is highly variable between the various shoulder movements. The measurement instruments and the methods that were used in previous studies were also different, as were the reliability statistics (such as the limits of agreement16). Previous publications have discussed the use of appropriate ICCs. Studies that have reported on the reliability of shoulder movements in healthy subjects16-22 may either overestimate or underestimate the reliability of shoulder movements when applied in a clinical setting. For reasons of clarity we provide the formulas of the ICCs that we used. ICC values are known to depend on the variation in the study group.16 More homogeneous study groups may produce lower ICCs as the main contributor of the variance (patient) is negligible. As the end point of range of movement was determined by the patients’ subjective reporting of pain the results of this study are only applicable to patients with shoulder pain. With further training, and with standardisation of methods of measurement, the rheumatologists may reach higher levels of reliability. Considering the variability in reliability, it may be questioned whether it is necessary to have patients perform all isolated shoulder movements, especially if there are time constraints. The scapular fixation may require extra training to perform in a standardised way. One way to overcome this would be to eliminate complex shoulder movements, and choose the most reliable movements—for example, hand behind back instead of internal rotation in abduction, which both assess internal rotation. Although we had the examiners do a single movement only, the ICCs may also improve with two readings.

In conclusion, we recommend that for clinical trials of painful shoulder disorders where shoulder movements are an end point the assessments should be performed by the same examiner whenever possible. The examiner requires training and practice in a standardised protocol. For rheumatologists, one hour’s training may not be enough. Consideration should be given to choosing only those measures with known reliability.

ACKNOWLEDGEMENTS

Measurement of shoulder movement

Electro-Med, Melbourne for the supply of the Plurimeter-V inclinometer.

Authors’ affiliations
J L Hoving, R Buchbinder, Department of Clinical Epidemiology, Cabrini Hospital, Australia
J L Hoving, R Buchbinder, A Forbes, Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia
Green, Monash Institute of Health Services Research, Monash Medical Centre, Clayton, Australia
N Bellamy, Centre of National Research on Disability and Rehabilitation Medicine, Royal Brisbane Hospital, Herston, Australia
C Brand, A Stockman, Centre for Rheumatic Disease, Royal Melbourne Hospital, Parkville, Australia
R Buchanan, Department of Rheumatology, Austin and Repatriation Medical Centre, Heidelberg, Australia
S Hall, Cabrini Hospital, Melbourne, Australia
M Patrick, Box Hill Hospital, Box Hill, Australia
P Ryan, Alfred Hospital, Prahran, Australia

REFERENCES
How reliably do rheumatologists measure shoulder movement?

J L Hoving, R Buchbinder, S Green, A Forbes, N Bellamy, C Brand, R Buchanan, S Hall, M Patrick, P Ryan and A Stockman

*Ann Rheum Dis* 2002 61: 612-616
doi: 10.1136/ard.61.7.612

Updated information and services can be found at:
[http://ard.bmj.com/content/61/7/612](http://ard.bmj.com/content/61/7/612)

These include:

**References**
This article cites 12 articles, 5 of which you can access for free at:
[http://ard.bmj.com/content/61/7/612#BIBL](http://ard.bmj.com/content/61/7/612#BIBL)

**Email alerting service**
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**
Articles on similar topics can be found in the following collections

- Degenerative joint disease (4641)
- Musculoskeletal syndromes (4951)
- Pain (neurology) (883)

To request permissions go to:
[http://group.bmj.com/group/rights-licensing/permissions](http://group.bmj.com/group/rights-licensing/permissions)

To order reprints go to:
[http://journals.bmj.com/cgi/reprintform](http://journals.bmj.com/cgi/reprintform)

To subscribe to BMJ go to:
[http://group.bmj.com/subscribe/](http://group.bmj.com/subscribe/)