Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols

Ann M. Cools, PT, PhD\textsuperscript{a,\ast}, Lieven De Wilde, MD, PhD\textsuperscript{b}, Alexander Van Tongel, MD\textsuperscript{b}, Charlotte Ceyssens, PT\textsuperscript{a}, Robin Ryckewaert, PT\textsuperscript{a}, Dirk C. Cambier, PT, PhD\textsuperscript{a}

\textsuperscript{a}Department of Rehabilitation Sciences and Physiotherapy, Faculty of Medicine and Health Sciences, University Hospital, Ghent, Belgium

\textsuperscript{b}Department of Orthopaedic Surgery and Traumatology, Faculty of Medicine and Health Sciences, University Hospital, Ghent, Belgium

Background: Shoulder range of motion (ROM) and strength measurements are imperative in the clinical assessment of the patient’s status and progression over time. The method and type of assessment varies among clinicians and institutions. No comprehensive study to date has examined the reliability of a variety of procedures based on different testing equipment and specific patient or shoulder position. The purpose of this study was to establish absolute and relative reliability for several procedures measuring the rotational shoulder ROM and strength into internal (IR) and external (ER) rotation strength.

Methods: Thirty healthy individuals (15 male, 15 female), with a mean age of 22.1 ± 1.4 years, were examined by 2 examiners who measured ROM with a goniometer and inclinometer and isometric strength with a hand-held dynamometer (HHD) in different patient and shoulder positions. Relative reliability was determined by intraclass correlation coefficients (ICC). Absolute reliability was quantified by standard error of measurement (SEM) and minimal detectable change (MDC). Systematic differences across trials or between testers, as well as differences among similar measurements under different testing circumstances, were analyzed with dependent \textit{t} tests or repeated-measures analysis of variance in case of 2 or more than 2 conditions, respectively.

Results: Reliability was good to excellent for IR and ER ROM and isometric strength measurements, regardless of patient or shoulder position or equipment used (ICC, 0.85-0.99). For some of the measurements, systematic differences were found across trials or between testers. The patient’s position and the equipment used resulted in different outcome measures.

Conclusions: All procedures examined showed acceptable reliability for clinical use. However, patient position and equipment might influence the results.

*Reprint requests: Ann M. Cools, PT, PhD, Department of Rehabilitation Sciences and Physiotherapy, University Hospital Ghent, De Pintelaan 185, 2B3, B-9000 Ghent, Belgium.

E-mail address: ann.cools@ugent.be (A.M. Cools).
Clinicians and researchers routinely evaluate changes in the status of patients over time. The assessment of range of motion (ROM) and muscle strength is important in (1) the diagnosis of glenohumeral disorders and pathologies, (2) the assessment of treatment progression and effectiveness, and (3) for quantifying the amount of change in movement quality and force development occurring over time. In addition, establishing objective measurements of ROM and strength is essential to the identification of risk factors for shoulder pain, particularly in an athletic population. It is therefore, important for clinicians and researchers to have accurate and reliable examination tools to objectively assess the functional status of the shoulder joint. In general, ROM and strength assessment are considered to be necessary outcome measures of shoulder function besides self-report outcome scores and subjective clinical examinations.

The method and type of functional shoulder assessment, including the patient’s position and testing equipment, varies among clinicians and institutions by factors such as time, the clinician’s educational background, availability of equipment, and the specific movement or muscle being assessed. Goniometry has been used widely for ROM assessment because of its low cost and portability. However, in the assessment of shoulder external (ER) and internal (IR) rotation ROM, the clinician is required to use both hands, which makes stabilization of the trunk and the scapula more difficult and thus often leads to increased risk for measurement errors or the need for a second assessor.

Inclinometry is another practical alternative in which gravity is used as a reference point for ROM measurements. Digital inclinometers are portable and lightweight but are more costly than goniometers. In addition, ROM measurements with an inclinometer can only be performed in a vertical plane because the tool depends on gravity for interpretation of ROM measurements.

A variety of methods have been used for the assessment of shoulder rotational strength, including manual muscle testing (MMT), hand-held dynamometry (HHD), and isokinetic testing. Although considered to be the gold standard, isokinetic testing is not always user-friendly because of the high costs and the laboratory setting required. HHD is a more objective evaluation method and far more superior to subjective MMT when evaluating changes in muscle strength after injury.

Numerous studies have examined the reliability of one specific testing protocol or novel equipment or have compared testing positions or equipment. In general, the following conclusions may be drawn: First, intra-rater and inter-rater reliability for the measurement of the passive movements of the shoulder varies with the method of measurement and the equipment used. Universal goniometers, as well as inclinometers, are recommended for the assessment of shoulder ROM. Standardized trunk and scapula fixation, as well as standardization of the amount of overpressure at the end ROM, increase reliability.

Second, considering HHD’s ease of use, portability, cost, and compact size, the HHD can be regarded as a reliable and valid instrument for shoulder muscle strength assessment in a clinical setting. However, results are prone to error that might arise from the strength of the investigator, the testing position, and the stabilization of the patient.

In clinical practice, the minimal detectable change (MDC) is one of the most important values to consider when using objective outcome measurements. The MDC is the minimum amount of change in a patient’s score that ensures the change is not the result of measurement error. The MDC is calculated in terms of confidence of prediction; for example, MDC is based on a 90% confidence interval. However, only a few studies have mentioned MDC results in the interpretation of their data regarding shoulder ROM and strength evaluation.

Although measurement techniques for shoulder rotation ROM and strength have been reported using supine, prone, and sitting procedures, as well as using varying shoulder positions from neutral rotational position up to 90° of abduction, to our knowledge, no study has combined all of these variables into one comprehensive reliability study performed by the same team of examiners. This condition mimics optimally the clinical reality in which often a team of health professionals—medical doctors and paramedic assistants—performs a set of tests using the available assessment tools. In addition, the study design allows statistical analysis including all measurements, and in particular, stating the MDC, which is known to be very relevant in clinical practice. Moreover, providing a variety of measurement protocols, based on established norms and potential functional requirements of the patient, this study may improve the quality of the patient’s assessment over time.

The purpose of this study was to examine the intra-rater and inter-rater reliability and the MDC for clinical goniometric, inclinometric, and HHD dynamometry measurements of shoulder passive ROM into ER and IR, and

**Keywords:** Shoulder rotation; range of motion measurement; strength measurement; reliability; goniometer; inclinometer; hand-held dynamometer
isometric strength into ER and IR, using different procedures, body and shoulder positions, and testing equipment. In addition, we hypothesized that ROM and strength measurements might be influenced by gender and the position of the patient and his or her shoulder.

**Materials and methods**

**Participants**

A convenient sample of 30 asymptomatic adults (15 women and 15 men) was recruited for this investigation from a local university setting during a 7-month interval. Estimated sample size was based on literature suggesting that with 2 ratsers, a significance level of 0.05 and a power of 80% to determine an ICC score of 0.7, that a minimum of 19 samples would be required. The mean (standard deviation) for the participants’ age, body weight, and height were 22.1 (1.4) years, 76.8 (17.8) kg, and 1.72 (1.9) meters, respectively. None of the participants reported a history of shoulder or neck pain or current participation in overhead sports on a competitive level. Inclusion and exclusion criteria were assessed with a questionnaire. Before participation, participants read and signed the informed consent form.

**Instruments**

Shoulder ROM was measured with 2 instruments: a plastic Baseline goniometer (Gymna hoofdzetel, Bilzen, Belgium), marked in 1° increments, with 2 adjustable overlapping arms, and an Acumar Digital Inclinometer (model ACU360; Lafayette Instrument Co, Lafayette, IN, USA). The manufacturer’s specifications indicate that this instrument is capable of measuring a range up to 180° with an accuracy of 1°.

Isometric muscle strength of the shoulder external and internal rotators was measured using a MicroFET 2 HHD (Hoggan Health Industries Inc., West Jordan, UT, USA).

All measurements were performed by 2 assessors who were familiarized with the procedures before the measurements. Examiner 1 was male, height, 1.73 meters; and weight, 75 kg; and examiner 2 was female, height, 1.63 meters; weight, 59.5 kg.

**Procedures**

After a standardized warm-up procedure consisting of multiplanar shoulder movements that was supervised by the testers, the testing procedure started. A summary of all measurements performed in this investigation is given in **Table I**. The procedure started with goniometric measurements (1-8) of the rotational ROM in different positions. Two examiners participated in these measurements: 1 examiner provided stabilization of the scapula and trunk and performed the shoulder ER or IR, and the other examiner handled the goniometer to perform the measurement. Subsequently, each examiner independently measured the shoulder in the various test positions (as mentioned in **Table I**) with the inclinometer for the ROM measurement and with the HHD for the strength measurement. Three 3 familiarization trials were performed for each measurement, and 3 subsequent trials were executed for further analysis. All measurements were performed in the order of (1) supine (measurements 12-14, 19-22), (2) sitting (measurements 9-11, 15-18), and (3) prone (measurements 23 and 24), thus randomizing ROM and strength measures and avoiding fatigue from the strength measurement. The raters were blinded to the results because the second assessor recorded the data from the first assessor and vice versa.

For the goniometric measurements, the goniometer was positioned with the fulcrum placed at the olecranon, the stable arm horizontal (1-4 and 8) or vertical (5-7), depending on the specific procedure, and on the moving arm along the forearm with the processus styloideus ulnae as a reference point.

For the ROM measurements with the inclinometer, the digital inclinometer was zeroed by using a fixed horizontal (measurements 9-11) or vertical (measurements 12 and 13) reference point, depending on the measurement, to ensure accuracy. Then, the inclinometer was placed dorsally or ventrally midway on the forearm, depending on movement direction.

For all ROM measurements, goniometric as well as inclinometric, scapular compensation was controlled by palpating the coracoid (with the examiner’s thumb) and the spine of the scapula (with the examiner’s fingers). This method has been shown to more accurately control for scapular movements compared with putting the full hand on the shoulder during the measurement.

For the strength measurements, the HHD was placed 2 cm proximal of the processus styloideus ulnae, on the dorsal (ER strength) or ventral (IR strength) forearm. Three repetitions of 5 seconds of maximal voluntary effort were performed using a “make” test (gradually increasing resistance up to maximum without “breaking” the subject’s strength). A 10-second resting period was given in between trials. Stabilization of the upper arm, shoulder, scapula, and trunk were provided by manual fixation by the examiner’s hand, arm, and trunk, if necessary.

**Statistical analysis**

Means (standard deviations) were calculated across participants for the dependent variables, the ROM values in degrees, and the absolute isometric strength data in Newton (N). All the dependent variables demonstrated a normal distribution (Kolmogorov-Smirnov test), and parametric tests were applied. To assess relative reliability—the degree to which individuals maintained their position in a sample with repeated measures—intraclass correlation coefficients (ICC) were calculated with the corresponding 95% confidence interval (CI). For the intra-rater reliability, the 3 trials for each measurement were used, and ICC_{3,k} (two-way random model—absolute agreement) was calculated because measurements were performed by 1 individual. For the inter-rater reliability, in which measurements from 2 independent assessors are compared to determine whether the particular instrument (in this case the inclinometer and the HHD) can be used with confidence and reliability among equally trained clinicians, the mean of the 3 trials were used, and ICC_{2,k} (two-way mixed model—absolute agreement) was calculated. Only intra-rater reliability was calculated for the goniometric measurements, and intra-rater as well as inter-rater reliability was determined for the inclinometer and HHD measurements.

ICC interpretation was based on the guidelines of Fleiss, according to whom high (>0.90) values reflect excellent reliability, values between 0.80 and 0.89, good; between 0.70 and 0.79, moderate; and values <0.70, low reliability. However, ICC values
may be influenced by intersubject variability of scores, and a large ICC may be reported despite poor trial-to-trial consistency if the intersubject variability is too high.21 Therefore, to examine the absolute reliability—the degree to which repeated measurements vary for individuals—the standard error of measurement (SEM) was calculated as $SD \times \sqrt{1 - ICC}$, where the SD is the SD from all scores from the participants.33 The SEM was used for calculating the minimal detectable change (MDC$_{90}$) which was calculated as $SEM \times \sqrt{2}$.

Paired samples t test (for inter-rater differences) and a general linear model analysis of variance (ANOVA) for repeated measures (for intra-rater variability) were used to examine whether there were systematic differences between testers or trials. In case of systematic differences among trials, a Bonferroni procedure was used as a post hoc test for pairwise comparisons.

In view of the research question of whether the gender, position of the patient, or the equipment used influences ROM or strength results, comparisons were performed between similar measurements with respect to measurement equipment (for ROM), gender, patient, and shoulder position. A general linear ANOVA for repeated measures with the within-subjects factor of “position,” and the between-subject factor of “gender” was used. In case of significant interaction or main effects in the ANOVA model, post hoc Bonferroni tests were performed. The $\alpha$ level was set at .05. All statistical analyses were performed using IBM SPSS 21 software (IBM Corp, Armonk, NY, USA).

## Results

Data from the reliability analysis for ROM measurements are summarized in supplementary Table I (available on the journal’s website at www.jshoulderelbow.org). ICCs for intra-rater reliability varied from 0.85 (IR in 90° forward flexion/goniometer/supine) to 0.99 (IR in 90° abduction/inclinometer/sitting) and for inter-rater reliability from 0.96 (IR in 90° forward flexion/sitting/inclinometer) to 0.99 (IR in 90° abduction/sitting/inclinometer), showing excellent intra-rater and inter-rater reliability of all procedures. The MDC$_{90}$, derived from the mean of 3 repetitions, varied from 4.02 N (IR in forward flexion/sitting/inclinometer) to 7.48 N (ER in 0° abduction). The MDC, derived from the mean of 3 repetitions, varied from 7.87 N (ER in

### Table 1  Overview of range of motion and HHD dynamometry measurements*

<table>
<thead>
<tr>
<th>Type of measurement</th>
<th>Testing equipment</th>
<th>Test position</th>
<th>Testing procedure</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROM</td>
<td>Goniometer</td>
<td>Sitting</td>
<td>ER at 0° abduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ER at 90° abduction</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IR at 90° abduction</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IR at 90° forward flexion</td>
<td>4</td>
</tr>
<tr>
<td>Supine</td>
<td></td>
<td>ER at 0° abduction</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ER at 90° abduction</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR at 90° abduction</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR at 90° forward flexion</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Inclinometer</td>
<td>Sitting</td>
<td>ER at 0° abduction</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ER at 90° abduction</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR at 90° abduction</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR at 90° forward flexion</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Supine</td>
<td></td>
<td>ER at 0° abduction</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ER at 90° abduction</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Isometric strength</td>
<td>HHD</td>
<td>Sitting</td>
<td>ER at 0° abduction</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IR at 0° abduction</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ER at 90°-90° abduction/ER</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IR at 90°-90° abduction/ER</td>
<td>18</td>
</tr>
<tr>
<td>Supine</td>
<td></td>
<td>ER at 0° abduction</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ER at 0° abduction</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ER at 90°-90° abduction/ER</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR at 90°-90° abduction/ER</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Prone</td>
<td></td>
<td>ER at 90°-90° abduction/ER</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IR at 90°-90° abduction/ER</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

* ER and IR measurements of ROM with the goniometer and the inclinometer, and isometric strength with the HHD, measured in a sitting, supine, or prone position, with the shoulder in 0° or 90° of abduction, 90°-90° abduction/ER, or 90° forward flexion. All measurements are numbered for further details in the text.
90° abduction/prone/examiner 1) to 22.11 N (IR in 0° abduction/sitting/examiner 2). Data are summarized in supplementary Table II (available on the journal’s website at www.jshoulderelbow.org).

From the $P$ values of the ANOVA for repeated measurements (between trials) and the dependent $t$ tests (between testers; see Supplementary Tables I and II), we can conclude that despite the high ICCs for all measurements, some are susceptible to systematic errors over trials and between testers. In particular, some ROM measurements with the inclinometer and some strength measurements in the supine and prone position seem to be prone to systematic errors.

Because the ANOVA for repeated measurements revealed no significant gender-based interaction or main effects, men and women were combined for further analysis. The results of the comparative analysis of similar measurements in different conditions according to testing equipment (goniometer vs inclinometer) and patient position (sitting vs supine) are summarized in Table II. The results show marked ROM differences between positions and equipment used, particularly for the measurement of IR in 90° of abduction. In addition, results from the goniometer and the inclinometer often differ significantly, despite the high reliability.

Results for strength measurements, comparing patient position (sitting, supine, and prone) and shoulder position (0° vs 90°-90° of abduction and ER) are described in Table III. The results show that also here there is no consistency in the outcome for most measurements comparing different measurement tools and patient positions. In particular, the prone position results in higher values for IR and lower values for ER strength compared with the other positions.

**Discussion**

This study established good to excellent reliability for IR and ER passive ROM measured with a goniometer or an inclinometer and isometric strength measurements using a HHD, regardless of the patient’s position, shoulder position, or equipment used. This is, to our knowledge, the first study to provide a comprehensive reliability analysis of ROM and strength measurements performed by the same examiners and combining several procedures frequently used in clinical practice.

**IR and ER ROM measurements**

When adhering to the procedures outlined in this investigation, goniometric and inclinometric measurements were found to be suitable instruments to quantify shoulder ROM, with ICCs from 0.85 to 0.99 and a slight advantage for the inclinometer of 0.89 to 0.99. Several studies, summarized in the systematic review by van de Pol et al,32 have examined the reliability of goniometry and digital
inclinometry separately to measure shoulder ROM, with considerably varying results.\(^{1,9,15,20,22}\) The trend for better reliability with an inclinometer was also observed in a study of Kolber et al,\(^{15}\) whereas Mullaney et al\(^{20}\) reported that reliability estimates were similar between the goniometer and the digital inclinometer. It should be noted that the Kolber study\(^{15}\) measured active ROM, whereas the measurement of interest in our study was the passive rotational ROM.

Moreover, contrary to these studies, goniometric measurements in our study were performed by 2 examiners. This probably increases patient stabilization and measurement accuracy, which is reflected in the high ICC values. The lowest ICC values for the goniometric measurements were in the IR ROM in 90° of forward flexion (0.85), with the patient sitting as well as supine. This might be due to difficult scapular stabilization in this position or to difficulties in defining the starting reference position. Indeed, some have suggested that manual scapular stabilization in an appropriate manner is important to better isolate glenohumeral IR but that ER is not much affected by manual stabilization.\(^{2,35}\)

For measurement error and what might constitute true change, the MDC\(_{90}\) for the intra-rater analysis indicated that a change from \(\geq 4.44^\circ\) to \(8.03^\circ\) (depending on the specific test) for the goniometer and from \(4.02^\circ\) to \(6.36^\circ\) for the inclinometer is required to be \(90\%\) certain that this change is not due to intra-rater variability of measurement error.\(^{5}\) The MDC\(_{90}\) for the inter-rater analysis of the inclinometric measurements revealed that a change from \(\geq 2.82^\circ\) to \(5.47^\circ\) (depending on the specific measurement) is required to be \(90\%\) certain that this change is not the result of inter-rater variability or measurement error. Only a few studies mention MDC results in the interpretation of their data, so we have little literature to compare our results with. Kolber et al\(^{16,17}\) found MDC values of \(8^\circ\) to \(9^\circ\) in their studies, which are slightly higher than in our study. However, the clinician should bear in mind that MDC values depend on the maximum value that can be obtained and that these differ according to the movement direction. In our case, the very low MDC\(_{90}\) of \(2.82^\circ\) for the measurement of IR supine into forward flexion is the result of the relatively small arc of motion of this movement, with a mean value of \(\pm 14^\circ\). To truly compare absolute reliability and measurement errors for different kinds of movements, the percentage SEM should be calculated. However, in view of the clinical perspective, we reported the MDC in the assumption that the MDC is the value the clinician uses in the interpretation of the measurement data obtained in a patient.

### Isometric strength measurement

Traditionally, isometric rotational shoulder strength is assessed in a seated,\(^{4,13,14,23}\) supine,\(^{7,10}\) or prone position.\(^{23}\) Our results regarding relative reliability reveal ICCs between 0.93 and 0.99, showing excellent reliability for all of the tests performed, regardless of the examiner, the position of the patient, or the shoulder position. Our results are slightly higher than the values previously reported.\(^{3,12,23}\) Although relative reliability values showed excellent reproducibility, MDC values still are an important issue to be considered. Depending on the specific measurement, MDC\(_{90}\) varied from 7.87 (ER prone 90°-90° examiner 1) to 26.60 (IR prone 90°-90° inter-rater analysis). Because other studies examining reliability of strength measurement procedures did not calculate MDC\(^{23}\) or used other statistical analysis, such as limits of agreement,\(^{4}\) we have no literature to compare our results with.

#### Systematic error of measurement between trials and testers

Despite the high ICCs for all measurements, our results revealed some systematic differences between trials and between testers for the ROM as well as for the strength measurements. In particular, some ROM measurements

### Table III Details of the comparative statistical analysis for isometric strength measurements based on position *

<table>
<thead>
<tr>
<th>Isometric strength</th>
<th>SI</th>
<th>SU</th>
<th>PR</th>
<th>Statistical test</th>
<th>ANOVA or dependent t test</th>
<th>Pairwise comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER in 0° abduction</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Dependent t test</td>
<td>&lt;.001(^1) NA</td>
<td></td>
</tr>
<tr>
<td>IR in 0° abduction</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Dependent t test</td>
<td>&lt;.001(^1) NA</td>
<td></td>
</tr>
<tr>
<td>ER in 90°-90° position</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>ANOVA repeated measurements with within subject factor “position” 3 levels</td>
<td>&lt;.001(^1) SI-SU: .624</td>
<td>SI-PR: .006(^1) SU-PR: &lt;.001(^1)</td>
</tr>
<tr>
<td>IR in 90°-90° position</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>ANOVA repeated measurements with within subject factor “position” 3 levels</td>
<td>.001(^1) SI-SU: .579</td>
<td>SI-PR: .042(^1) SU-PR: &lt;.001(^1)</td>
</tr>
</tbody>
</table>

*P* values for the ANOVA for repeated measurements or *t* tests and for post hoc pairwise comparisons between conditions, with NA indicating no post hoc test performed.

\(^1\) *P* < .05, indicating statistical significance.
with the inclinometer and some strength measurements in the supine and prone position seem to be prone to systematic errors. Systematic differences in the strength measurements might be the result of differences in the strength of the testers.

**Comparison of similar measurements using different equipment or positions, or both**

Although not the primary research question but relevant for clinicians, the differences between procedures regarding a similar measurement was performed. Several significant differences became apparent, based on equipment and on patient and shoulder position. In particular, ROM into IR lacks consistency. For the strength measurements, the prone position seems to correlate the least with the other positions tested in this study. In conclusion, the clinician should remember that, despite the high relative and absolute reliability, the procedures should not be mixed in the evaluation of 1 patient in view of the systematic differences found between procedures and positions. We recommend that clinicians use at least 2 positions to establish the strength impairment in a more functional position for the patient but urge the examiner to take into account the possible differences in outcome by patient and shoulder position.

**Limitations and future research**

Some limitations of this study need to be noted. Owing to technical issues, it was not possible to perform some testing (for instance because of the limitations of the use of the inclinometer depending on gravity for interpretation of movement angles) or statistical analysis (inter-rater reliability analysis with the goniometer because both assessors were needed during the protocol). In addition, despite extreme effort to standardize not only the measurement but also the stabilization of the patient’s trunk and scapula, no external fixation was used for reasons of clinical relevance. Clinicians in the field do not always have access to external devices to measure ROM and strength, and very often, external fixation makes the procedure more time-consuming and, thus, less attractive for the clinician, but might have influenced our results. The end range during the ROM measurements was determined by subjective criteria, based on familiarization trials by the assessors, but not objectively controlled. An inclinometer was recently introduced that measures simultaneous angle (in degrees) and force (in kg), making it possible to quantify the load applied at the end of ROM during passive movement testing. Further studies should consider the use of these kind of devices for ROM measurement; however, further proof of reliability and validity is needed.

Finally, the use of asymptomatic individuals needs to be acknowledged as a limitation. Future studies should examine not only the reliability but also the responsiveness of these procedures in patients with various pathologic shoulder conditions.

**Clinical implications**

Our study shows excellent relative reliability for several protocols for the measurement of ROM and strength into ER and IR of the shoulder and clinically acceptable absolute reliability values. Our results show that none of the procedures examined should be avoided, and it is up to the clinician’s preference to use a specific examination protocol. From the perspective of practical utility, we recommend the supine position for all ROM measurements because maximal trunk and scapula stabilization are assured in this position.

With respect to testing equipment, the goniometer and the inclinometer show excellent reliability. The goniometer shows slightly higher accuracy and the inclinometer has the advantage that only 1 tester is needed to perform the testing.

For the strength measurements, sitting and supine positions both seem to give a reliable evaluation of muscle function, whereas the prone position should be avoided, in view of the large differences with the strength results in other positions.

**Conclusions**

The purpose of this study was to establish absolute and relative reliability for several procedures measuring the rotational shoulder ROM and strength into IR and ER. The study results show good to excellent reliability values for all procedures performed. Clinicians should consider their choice based on the available equipment and the ability of the patient to achieve the body or shoulder position. In general, measurements in the supine position are recommended because of practical applicability and body stabilization, and clinicians are recommended to use more than 1 procedure to allow functional measurements based on the patient’s abilities at the moment of evaluation.

**Supplementary data**

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jse.2014.01.006.

**References**


30. van de Pol RJ, van Trijffel E, Lucas C. Inter-rater reliability for measurement of passive physiological range of motion of upper extremity joints is better if instruments are used: a systematic review. J Physiother 2010;56:7-17.

