Technical and measurement report

Intra- and inter-rater reliability of a modified measure of hand behind back range of motion

Paul A. van den Dolder a, b, *, Paulo H. Ferreira a, Kathryn Refshauge a

a Discipline of Physiotherapy, Faculty of Health Sciences, The University of Sydney, Sydney, Australia
b Illawarra Shoalhaven Local Health District, Wollongong, Australia

A R T I C L E   I N F O

Article history:
Received 14 April 2013
Received in revised form
30 July 2013
Accepted 22 August 2013

Keywords:
Reproducibility of results
Range of motion, articular
Shoulder

A B S T R A C T

The aim of this reliability study was to identify the clinimetric properties, specifically intra- and inter-rater reliability, for measuring the functionally and clinically important hand behind back (combined shoulder internal rotation/adduction and elbow flexion) range of motion using a modified technique. Sixty asymptomatic participants (20 male, 40 female) aged 45.4 ± 11.7 years (mean ± SD). Hand behind back was measured as the distance from the mid-line between the posterior superior iliac spines (PSIS) to the tip of the thumb and recorded in centimetres above the PSIS (a positive measure) or below the PSIS (a negative measure). The intra- and inter-rater reliability for the measurement were excellent, with ICC (2,1) of 0.95 for intra-rater and ICC (2,2) of 0.96 for inter-rater reliability. The standard error of the measurement (SEM) and smallest real difference (SRD) values showed acceptable levels of measurement error, SEM 4.3 mm and SRD 12.8 mm for intra-rater reliability and SEM 2.6 mm and SRD 7.7 mm for inter-rater reliability. The assessment of hand-behind-back motion from a point equidistant between the PSISs to the tip of the thumb is more reliable than traditional methods and a useful clinical tool.

1. Introduction

In order to justify intervention and judge its effectiveness, clinicians need to use reliable measurement methods in their clinical practice. Many measurements may be taken by the same or by different clinicians in the management of one patient. It is therefore essential to understand whether the measurements are reliable both within and between clinicians. Adequate hand behind back (consisting of combined shoulder internal rotation/adduction and elbow flexion) range of motion in the shoulder is fundamental to the normal performance of many useful activities of daily living including dressing, washing and using the arm whilst rising from a chair (Hertling and Randolph, 2006). Restriction of this movement is often seen post-injury or post-immobilisation (Rockwood et al., 1998). The assessment of hand behind back range is therefore considered important for restoration of function and is routinely assessed in shoulder examinations (Hertling and Randolph, 2006).

Various methods have been investigated for measuring the hand behind back movement. These include estimation of the distance between the tip of the thumb behind the back and a specific vertebral level (inter-rater ICC range 0.12–0.80; intra-rater ICC range 0.016–0.91) (Hayes et al., 2001; Edwards et al., 2002; Hoving et al., 2002); tape measure from thumb tip to T1 (inter-rater reliability ICC 0.39; intra-rater reliability ICC 0.39); tape measure from T1 to ulnar styloid process (inter-rater reliability ICC 0.71 to 0.75; intra-rater reliability ICC 0.84) (Green et al., 1998) and estimation on a 6 point scale of the landmark reached (inter-rater agreement 13%; intra-rater agreement 33%) (Bostrom et al., 1991). These wide variations in reliability have been attributed to the complexity of the movement itself, the possible attenuated range of available scores in different techniques (Hayes et al., 2001), and the difficulty in reliably palpating anatomical landmarks for techniques (Edwards et al., 2002). In particular, palpation of vertebral levels has been shown to have poor inter- and intra-rater reliability (Christensen et al., 2002; T1–T8 Inter-rater Kappa 0.22–0.24, Intra-rater Kappa = 0.59–0.64; Mootz et al., 1989; L1 – S1 Inter-rater Kappa = −0.09 to 0.48, Intra-rater Kappa = −0.17 to 0.17). In contrast the palpation of PSIS has higher reliability (Wiangkham et al., 2010: Intra-rater Kappa = 0.94; Simmonds and Kumar, 1993: within rater error = 8 mm, between rater error = 20 mm).

Another common method (modified by van den Dolder and Roberts, 2003) used by clinicians to assess hand-behind-back range of motion is to measure the distance between a point
equidistant between the posterior superior iliac spines (PSIS) of the pelvis to the tip of the thumb reaching up the spine. This method has the potential advantage of utilising the PSIS, which are easier to palpate than vertebral bodies (Simmonds and Kumar, 1993; Wiangkham et al., 2010). The purpose of this study therefore, was to examine the inter-rater and intra-rater reliability of this method.

2. Methods

Ethical approval was received from the University of Sydney. Written consent was obtained from all participants prior to entry to the study.

2.1. Participants

Sixty (20 male; 40 female) asymptomatic participants volunteered for this study and provided informed consent. To be included in the study volunteers had to be aged between 18 and 80 years and able to understand spoken English. Participants were excluded from the study if they had current shoulder pain, a history of trauma to the shoulder region in the previous four weeks, or if they had inflammatory arthritis or other serious pathology of the shoulder. The aim was to recruit a heterogeneous sample of convenience with a cross-section of different ages, body types and body flexibility levels.

2.2. Raters

Two raters (Rater A and Rater B) were used in this study. Both were experienced physiotherapists with over fifteen years clinical experience each in musculoskeletal physiotherapy. A standardised protocol for the measurement of hand-behind-back motion was followed:

- Raters located participants’ posterior superior iliac spines (PSIS);
- A horizontal line was marked to join the PSIS;
- The participant was then asked to reach the thumb up their spine as far as they could comfortably reach;
- The distance between the tip of their thumb and the mid-line between the two PSISs was then measured using a tape measure.
- Results were recorded in centimetres above the line (a positive measure) or below the line (a negative measure) as shown in Fig. 1.

During the movement the participant was asked to maintain their usual standing posture.

2.3. Randomisation

A random number generator was used to randomise order of test arm and order of raters. Numbers were sealed in opaque envelopes which were opened by the participant.

2.4. Inter-rater reliability

The same hand behind back measure was taken by both raters immediately after each other. Raters were in separate rooms so that...

Fig. 1. Measurement of hand behind back using tape measure to measure distance from PSIS level (point A) to participant’s tip of thumb (point B).

Fig. 2. Flowchart of progress of participants through the trial.
they did not have any contact with each other during the measurement procedure. Any marks made during the procedure were carefully removed so that they were not visible to the other rater. Results were recorded separately by each rater so that results were not available to the other rater.

2.5. Intra-rater reliability

Rater A repeated the measurement of hand-behind-back range two weeks after the initial measurement. Measurements were recorded in a separate booklet. The results of the initial measurement were not available to the rater.

2.6. Statistical analysis

To establish reliability, intra-class correlations (ICCs) were calculated based on the method of Shrout and Fleiss (1979). For intra-rater reliability an ICC (2,1) was performed and for inter-rater reliability an ICC (2,2) was used (Portney and Watkins, 2009). Cut-off values for acceptable reliability were based on those adopted by Rheault et al. (1992) where ICC > 0.8 was used to indicate high reliability; ICC > 0.6 to < 0.8 moderate reliability; ICC > 0.4 to < 0.6 fair reliability; and ICC ≤ 0.4 poor reliability.

The Standard Error of the Measurement (SEM), defined as the square root of the error variance, was used to quantify the test–retest reproducibility of results as described by de Vet et al. (2006). To estimate the smallest detectable measurement change beyond reproducibility noise the Smallest Real Difference (SRD) method was used and expressed as both a measure in centimetres and as a percentage of the measurement (Beckerman et al., 2001). As no benchmark for the acceptable SRD exists the authors used their clinical impression to determine whether the SRD observed was at an acceptable level for use. Bland–Altman plots were used to plot the difference scores against the mean for each participant (Bland and Altman, 1986).

2.7. Sample size

Sample size was determined using the method of Walter et al. (1998). We aimed for reliability of at least ICC = 0.8, and a priori determined that reliability ICC of 0.6 would be the least acceptable reliability measurement. Hence we defined our null hypothesis of ICC > 0.6 and alternate hypothesis of ICC > 0.8. Using the tables provided by Walter et al. (1998) assuming a 5% level of significance, power of 0.8 and a dropout rate of 20% a sample size of 60 subjects was required.

3. Results

A flow chart of participants in the trial is shown in Fig. 2. A summary of participants’ characteristics is shown in Table 1.

The intra- and inter-rater reliability for the measurement was excellent, as evidenced by ICC (2,1) of 0.95 for intra-rater reliability and 0.96 for inter-rater reliability (Table 2). The standard error of the measurement and smallest real difference values showed the measurement error to be acceptable; SEM 4.3 mm (1.27%) for intra-rater reliability and SEM 2.6 mm (0.76%) for inter-rater (Table 2).

Bland–Altman plots indicated that there was no systematic measurement bias for either intra-rater (Fig. 3) or inter-rater measurements (Fig. 4).

4. Discussion

The reliability of measurements used in the clinic or in research must be established to confidently interpret results that are meaningful on different testing occasions. In the current study, a modified measurement of hand behind back movement from the tip of the thumb to a point between the PSIS was found to have excellent inter- and intra-rater reliability properties, exhibiting better reliability scores than previously examined methods. The SRD values were low for both inter-rater and intra-rater reliability, indicating that this method of measuring hand-behind-back movement is sensitive to small changes; 0.77 cm for between-rater measurements and 1.28 cm for repeated measurements by the same rater. The higher SRD values for intra-rater measures could reflect within-subject variation from the time lapse of two weeks between measurements.

The results from our study also indicate a wide variation in hand-behind-back ranges in an asymptomatic population. It is currently unclear what range of motion for the hand-behind-back movement is considered within normal limits, or the range required to perform activities of daily living. In addition, the minimally clinically important difference in hand-behind-back measures is unknown for patients with shoulder disorders. Further investigation in these areas would be an important addition to understanding more about this functional movement and how much change is considered meaningful to patients with shoulder disorders.

The results of this study indicate that this method of measuring hand behind back movement has greater intra and inter-tester reliability than any other previous method studied. Other methods previously studied relied on the estimation of the vertebral level reached by the rater’s thumb, or measurement of the distance between the thumb or the ulnar styloid process and vertebral level reached (Bostrom et al., 1991; Green et al., 1998; Hayes et al., 2001; Edwards et al., 2002; Hoving et al., 2002). It should not be surprising that measurement methods that rely on

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Participants’ characteristics and hand behind back measures. Data are mean (SD) unless otherwise indicated.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Mean</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>45.8 (12.1)</td>
</tr>
<tr>
<td>Males:Females (n--)</td>
<td>20:40</td>
</tr>
<tr>
<td>HBB initial score (cm) – Rater A</td>
<td>34.4 (5.1)</td>
</tr>
<tr>
<td>HBB initial score (cm) – Rater B</td>
<td>33.6 (5.5)</td>
</tr>
<tr>
<td>HBB follow-up score (cm) – Rater A</td>
<td>34.3 (5.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Intra-rater and inter-rater reliability.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intra-rater reliability</td>
<td>Inter-rater reliability</td>
</tr>
<tr>
<td>ICC 2,1 (95% CI)</td>
<td>SEM (cm)</td>
</tr>
<tr>
<td>Hand-behind-back measure</td>
<td>0.95 (0.91—0.97)</td>
</tr>
<tr>
<td>SEM (cm)</td>
<td>SEM (%)</td>
</tr>
<tr>
<td>0.96 (0.93—0.98)</td>
<td>0.26</td>
</tr>
</tbody>
</table>

ICC = Intra-class correlation; SEM = standard error of the measurement; SRD = smallest real difference.
the unreliable palpation of spinal level will also tend to be unreliable. The method examined in this study utilises the more reliable palpation of the PSISs and provides for a greater range of scores than methods where vertebral level reached is scored. We found our modified method of measuring hand behind back motion is highly reliable, and recommend its use in the research setting.

4.1. Limitations

This study was conducted on an asymptomatic cohort. The results could be different in a symptomatic population where there is greater limitation in hand-behind-back range of motion, especially where patients are unable to reach the centre of their back. In addition, there were only two raters included in this trial. A better understanding of the range of intra- and inter-rater correlations would be achieved through using a larger group of assessors, including those with varying degrees of experience.

Financial disclosure statement

We certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on us or on any organisation with which we are associated and we certify that all financial and material support for this research and work are clearly identified in the title page of the manuscript.

This material has not been previously presented.
This trial received no financial assistance.

Acknowledgements

The authors gratefully acknowledge the assistance of Kristina van den Dolder for her invaluable assistance in collecting the data for this study.

References


