Interrater and intrarater reliability of the pectoralis minor muscle length measurement in subjects with and without shoulder impingement symptoms

Filip Struyfa,b,*, Mira Meeusa,b,c, Erik Fransend, Nathalie Roussel a,b, Nick Jansen e, Steven Truijena,Jo Nijs b,f

aDepartment of Rehabilitation Sciences and Physiotherapy, Faculty of Medicine and Health Sciences, University of Antwerp, Belgium
bChronic Pain and Chronic Fatigue Research Group (Pain in Motion)
cRehabilitation Sciences and Physiotherapy, Ghent University, Ghent, Belgium
dStatUa Center for Statistics, University of Antwerp, Belgium
eAZ Monica, Deurne, Belgium
fDepartment of Physical Medicine and Physiotherapy, University Hospital Brussels, Brussels, Belgium

ABSTRACT

Measuring the pectoralis minor muscle length (PML) is of clinical interest, as a short PML has been associated with a decrease of scapular posterior tilting and shoulder pain. However, as no reliability data are available at present, the objective of this study was to examine the inter- and intrarater reliability of the PML measurement in both subjects with and without shoulder impingement symptoms (SIS). Therefore, two assessors performed the PML measurement (3 times/shoulder) in 25 patients with SIS and 25 pain-free controls. Both assessors were blinded for each other’s findings. For reliability testing, intraclass coefficients (ICCs; model 2,1) and standard errors of measurements were calculated. Intrarater reliability analysis resulted with ICCs ranging from 0.87 (Standard error of measurement (SEM) 0.21–0.27%) (symptomatic) to 0.93 (SEM 0.19–0.30%) (asymptomatic) in patients with SIS, representing excellent test-retest agreement. Healthy subjects presented with ICCs ranging from 0.76 (SEM 0.29–0.32%) (dominant side) to 0.87 (SEM 0.21–0.32%) (non-dominant side), representing good test-retest agreement. ICCs and SEMs on the symptomatic and asymptomatic side (0.48 and 0.46%; 0.56 and 0.61%) in SIS patients, and on the two sides (non-dominant: 0.47 and 0.45%, dominant: 0.53 and 0.38%) respectively in healthy subjects showed moderate interrater reliability and low dispersion of the measurement errors. We concluded that the PML measurement has good to excellent intrarater reliability and poor to moderate interrater reliability.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Shoulder pain is a highly reported musculoskeletal disorder (Luime et al., 2004), in which up to 40% suffer from shoulder impingement symptoms (SIS) (van der Windt et al., 1995; Luime et al., 2004; Lewis, 2009). A variety of mechanisms that relate SIS to posture are already described in the literature, such as forward head posture, scapular protraction, humeral internal rotation and increased thoracic kyphosis (Borstad, 2006). Scapular protraction is defined as scapular anterior tilting together with scapular internal rotation. This deviation is proposed to reduce the subacromial area and has regularly been associated with SIS (Struyf et al., 2011). Measuring scapular protraction as the distance between the most posterior border of the acromion and the examining table has previously demonstrated to be a reliable tool for clinical practice (Nijs et al., 2005; Lewis and Valentine, 2007; Struyf et al., 2009). However, due to its lack of specificity, a direct measurement of the pectoralis minor muscle (PM) length was proposed (Lewis and Valentine, 2007; Struyf et al., 2012). Tate et al. proposed an alternative method for assessing the PML (Tate et al., 2012). Although they used the same reference points as in the present study, they normalized the PML through clavicular length instead of body length.
Disability Questionnaire (SDQ), and at least 2 out of 3 impingement tests had to be positive for inclusion (Hawkins test, Neer test & Jobe test were performed). Healthy controls were recruited by the following inclusion criteria: (1) age over 18, (2) the absence of shoulder pain (<5% on VAS or SDQ) during the last year and (3) the absence of a history of fractures, treatment or surgery of the shoulder girdle. General exclusion criteria were: (1) the inability to communicate using the Dutch language, (2) presence of a systemic disease, influencing shoulder pathology, (3) pregnancy, (4) traumatic shoulder pathologies and (5) shoulder pain radiating from the cervical spine.

All subjects were informed about the purpose and design of the study and gave their written informed consent. Subjects were free to withdraw from the study at any time. The study protocol was reviewed and approved by the Medical Ethics Committee of the University Hospital Brussels (ref: BUN B143201214180).

2.2. Research design

Prior to the study, 2 raters (JG; FN; Bachelors in physiotherapy with 1 year of clinical experience) underwent a 2-hours training session. This session was supervised by 1 physiotherapist with 9 years of clinical experience (FS; Master and Doctoral degree in sports physiotherapy). The training session instructed the raters in performing an accurate measurement of the PML, including a pilot testing on 8 healthy subjects (not included in data analysis).

First, the subjects’ body weight and body height were measured. Next, both shoulders were measured 3 times by each rater. The order of testing (between all 6 measurements and for the choice of assessor) was randomized by coin toss. First the order of rater was tossed, next the order of sides. Successively, both shoulders were tested independently. Both raters were blinded for hand dominance, the affected side (in case of shoulder pain patients), SDQ and pain-scores, and to the outcome of each other’s findings. In order to assure blinding, each assessor left the examining room when the other assessor performed an assessment. In order to detect the exact location of the coracoid process and rib 3,4 & 5, palpation was used.

2.3. PML measurement procedure

For the anthropometric measurement of the PML, we adapted the protocol of Borstad (2008), as previously described by Cool et al., (2010). The instruction was to palpate two anatomical reference points which in line represent the PML: (1) the inferomedial aspect of the coracoid process and (2) the caudal edge of the fourth rib at the sternum (Fig. 1). The distance between these two bony reference points was measured with a Vernier caliper (Hogetex, Varsseveld, The Netherlands, 0–300 mm). Both assessors were
blinded from the results on the caliper, until after the measurement. This was achieved by turning the caliper such that the reading scale was directed towards the subject. In order to neutralize variations of muscle length resulting from respiration, subjects were asked to exhale before the measurement and to inhale only after the measurement. The subjects were in supine position, with the elbows extended alongside the body with the palm placed on the examining table to minimize postural influences of the thoracic spine and to optimize the muscle relaxation of the surrounding musculature (Cools et al., 2010). By creating a Pectoralis Minor Index (PMI = (PML (cm)/subjects’ height (cm)) *100), the subjects’ variability in body height was normalized (Borstad and Ludewig, 2005) and the PML was expressed as a percentage of the subjects height. Between every measurement, each subject was asked to stand up before continuing the next measurement.

2.4. Pain and disability assessment

In order to evaluate self-reported shoulder disability, all patients completed the Shoulder Disability Questionnaire (SDQ) (van der Windt et al., 1998). In 16 items the patient described a possible pain-provocation during the last 24 h. Completion took between 5 and 10 min. The score was calculated by the summation of all yes-answers, divided by all answered questions (yes or no) and subsequently multiplied by 100. This resulted in a score between 0 (no disabilities) to 100 (severe disabled). The SDQ-NL (Dutch version) is suggested to be responsive and ready for use in clinical trials and longitudinal studies (van der Windt et al., 1998; van der Heijden et al., 2000; Paul et al., 2004).

Severity of shoulder pain was assessed with a Visual Analogue Scale (VAS-100 mm), a pain scoring system that is believed to be reliable, valid, and sensitive to change (Jensen et al., 1986). This pain assessment was performed both at rest and during shoulder activity.

2.5. Statistical analysis

The Statistical Package for the Social Sciences was used for analyzing the collected data (version 12.0 for Windows; SPSS Inc. Chicago, IL). Normality of the variables was visually tested for a Gaussian distribution and additionally tested with a 1-sample Kolmogorov–Smirnov goodness-of-fit test (p > 0.05; data not shown). The differences between both raters were calculated with a paired t-test. First, repeated measures ANOVA was performed in order to report the variance within subjects and to calculate the Standard Error of Measurement (SEM) (Weir, 2005). Next, as both raters tested all subjects and the analysis can be used to generalize to other levels (i.e. including systematic error), a 2,1 ICC model or 2-way random-effects ICC model was used (absolute agreement) (Shrout and Fleiss, 1979; Weir, 2005). For the healthy subjects, the reliability of the dominant and non-dominant shoulder was calculated. For shoulder SIS patients, the symptomatic and asymptomatic shoulder was used. Hereby a reliability coefficient less than 0.50 was an indication of poor reliability; reliability coefficients between 0.50 and 0.75, an indication of moderate reliability; reliability coefficients between 0.76 and 0.90, an indication of good reliability; reliability coefficients over 0.90 an excellent reliability (Portney and Watkins, 2000). The minimal detectable change with 95% confidence bounds (MDC95) \[ \text{MDC95} = \frac{\text{SEM}}{\sqrt{2}} \] was calculated (McKenna et al., 2004).

3. Results

The mean pain score at rest was 25.7 ± 25.3 mm (VAS) and 58.8 ± 18.3 mm during shoulder movements. Table 2 presents the PMI results of subjects with and without SIS, together with the

Table 2

<table>
<thead>
<tr>
<th>Subjects without shoulder impingement symptoms (n = 25)</th>
<th>Rater 1 (JG): mean PMI in % (SD)</th>
<th>Rater 2 (FN): mean PMI in % (SD)</th>
<th>Difference between raters (p value)</th>
<th>ICC*</th>
<th>ICC**</th>
<th>SEM (PMI)</th>
<th>MDC95 (PMI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant side</td>
<td>9.17 (0.54)</td>
<td>8.86 (0.65)</td>
<td>0.009</td>
<td>0.53</td>
<td>0.67</td>
<td>0.38</td>
<td>1.07</td>
</tr>
<tr>
<td>Non-dominant side</td>
<td>9.22 (0.39)</td>
<td>9.21 (0.62)</td>
<td>0.932</td>
<td>0.47</td>
<td>0.64</td>
<td>0.45</td>
<td>1.24</td>
</tr>
<tr>
<td>Symptomatic side</td>
<td>9.66 (0.68)</td>
<td>9.27 (0.69)</td>
<td>0.007</td>
<td>0.48</td>
<td>0.65</td>
<td>0.46</td>
<td>1.29</td>
</tr>
<tr>
<td>Asymptomatic side</td>
<td>9.64 (0.72)</td>
<td>9.52 (1.07)</td>
<td>0.517</td>
<td>0.56</td>
<td>0.72</td>
<td>0.61</td>
<td>1.68</td>
</tr>
</tbody>
</table>

2.4. Pain and disability assessment

In order to evaluate self-reported shoulder disability, all patients completed the Shoulder Disability Questionnaire (SDQ) (van der Windt et al., 1998). In 16 items the patient described a possible pain-provocation during the last 24 h. Completion took between 5 and 10 min. The score was calculated by the summation of all yes-answers, divided by all answered questions (yes or no) and subsequently multiplied by 100. This resulted in a score between 0 (no disabilities) to 100 (severe disabled). The SDQ-NL (Dutch version) is suggested to be responsive and ready for use in clinical trials and longitudinal studies (van der Windt et al., 1998; van der Heijden et al., 2000; Paul et al., 2004).

Severity of shoulder pain was assessed with a Visual Analogue Scale (VAS-100 mm), a pain scoring system that is believed to be reliable, valid, and sensitive to change (Jensen et al., 1986). This pain assessment was performed both at rest and during shoulder activity.

2.5. Statistical analysis

The Statistical Package for the Social Sciences was used for analyzing the collected data (version 12.0 for Windows; SPSS Inc. Chicago, IL). Normality of the variables was visually tested for a Gaussian distribution and additionally tested with a 1-sample Kolmogorov–Smirnov goodness-of-fit test (p > 0.05; data not shown). The differences between both raters were calculated with a paired t-test. First, repeated measures ANOVA was performed in order to report the variance within subjects and to calculate the Standard Error of Measurement (SEM) (Weir, 2005). Next, as both raters tested all subjects and the analysis can be used to generalize to other levels (i.e. including systematic error), a 2,1 ICC model or 2-way random-effects ICC model was used (absolute agreement) (Shrout and Fleiss, 1979; Weir, 2005). For the healthy subjects, the reliability of the dominant and non-dominant shoulder was calculated. For shoulder SIS patients, the symptomatic and asymptomatic shoulder was used. Hereby a reliability coefficient less than 0.50 was an indication of poor reliability; reliability coefficients between 0.50 and 0.75, an indication of moderate reliability; reliability coefficients between 0.76 and 0.90, an indication of good reliability; reliability coefficients over 0.90 an excellent reliability (Portney and Watkins, 2000). The minimal detectable change with 95% confidence bounds (MDC95) \[ \text{MDC95} = \frac{\text{SEM}}{\sqrt{2}} \] was calculated (McKenna et al., 2004).

3. Results

The mean pain score at rest was 25.7 ± 25.3 mm (VAS) and 58.8 ± 18.3 mm during shoulder movements. Table 2 presents the PMI results of subjects with and without SIS, together with the
shortening of this muscle as a cause or contribution to their symptoms (Lewis and Valentine, 2007). In addition, the acromion-
to-the-table measurement was not related to body length. When data are related to body length, scores are likely to reduce in var-
ance and therefore possible reduce reliability. However, the mea-
surement technique proposed in the current study achieved lower intra-
rater reliability data than the Lewis & Valentine’s study and its
diagnostic accuracy still needs to be studied. In addition, the largest
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.

It was shown that the intrarater ICC values were higher in
symptomatic patients. There was the possibility of a learning effect,
since individuals without symptoms were measured during the
first period of the study. The difficulty of the PML measurement
involves the precision in which bony landmarks are palpated. In
order to flatten out the learning curve, a training session longer
than the used 2-h training session could be encouraged. Second, the
low variability in the PML in the healthy controls, could explain
these lower ICC values. Finally, it should be mentioned that both
raters were aware that their judgments will be compared with
those of the other rater, increasing the possibility that the rater’s
behavior is altered because of the awareness of being observed, also
described as the Hawthorne effect (Kottner et al., 2011).

In conclusion, the findings of this investigation suggest that
intrarater reliability analysis resulted with ICCs ranging from 0.87
(0.65 and 0.46%; 0.72 and 0.61%) in SIS patients, and on the two
asymptomatic side in SIS patients; within 1 rater) was 0.49%. However,
difference in PMI between symptomatic and all asymptomatic sides
(both dominant & non-dominant in healthy subjects and asympto-
matic side in SIS patients; within 1 rater) was 0.49%. However,
with SEMs up to 0.61%, it is not likely that we will be able to
differentiate between patients and pain free subjects.