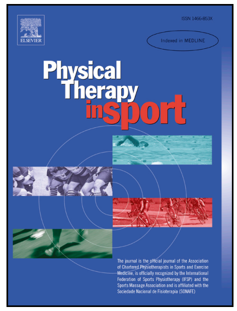
RESEARCH REPOSITORY

This is the author's final version of the work, as accepted for publication following peer review but without the publisher's layout or pagination. The definitive version is available at:

https://doi.org/10.1016/j.ptsp.2018.04.005

Møller, M., Attermann, J., Myklebust, G., Lind, M., Sørensen, H., Hebert, J.J., Nielsen, R.O., Bertelsen, S.L. and Wedderkopp, N. (2018) The inter- and intrarater reliability and agreement for field-based assessment of scapular control, shoulder range of motion, and shoulder isometric strength in elite adolescent athletes. Physical Therapy in Sport

Accepted Manuscript



The inter- and intrarater reliability and agreement for field-based assessment of scapular control, shoulder range of motion, and shoulder isometric strength in elite adolescent athletes

M. Møller, J. Attermann, G. Myklebust, M. Lind, H. Sørensen, J.J. Hebert, R.O. Nielsen, S.L. Bertelsen, N. Wedderkopp

PII: S1466-853X(17)30294-8

DOI: 10.1016/j.ptsp.2018.04.005

Reference: YPTSP 874

To appear in: *Physical Therapy in Sport*

Received Date: 7 July 2017

Revised Date: 2 March 2018

Accepted Date: 4 April 2018

TITLE PAGE

The inter- and intrarater reliability and agreement for field-based assessment of scapular control, shoulder range of motion, and shoulder isometric strength in elite adolescent athletes

M. Møller (PT, PhD) (1), J. Attermann (PhD) (2), G. Myklebust (PT, Professor) (3), M. Lind (Professor) (4), H. Sørensen (PhD) (1), J. J. Hebert (DC, PHD) (5,6), R. O. Nielsen (PT, PhD) (1), S. L. Bertelsen (PT) (7), N. Wedderkopp (Professor) (8)

(1) Department of Public Health, Section of Sport Science, Aarhus University, Aarhus, Denmark.

(2) Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark

(3) Oslo Sports Trauma Research Centre, Norwegian School of Sport Sciences, Oslo, Norway

(4) Div. of Sportstraumatology, Aarhus University Hospital, Aarhus, Denmark.

(5) School of Psychology and Exercise Science, Murdoch University, Murdoch, Western Australia, Australia

(6) Faculty of Kinesiology, University of New Brunswick, Fredericton, New Brunswick, Canada

(7) VIA University College, Aarhus, Denmark

(8) Sport Medicine Clinic, Orthopaedic dep. Hospital of Lillebaelt, Institute of Regional Health Service Research and Centre for Research in Childhood Health, IOB, University of Southern Denmark, Odense, Denmark.

Corresponding Author: Merete Møller, PT, PhD, Department of Public Health, Section of Sport Science, Aarhus University, 8000 Aarhus C, Denmark. Tel: +45 23367027, Fax: None, E-mail: *memoller@health.sdu.dk*

ACKNOWLEDGEMENTS

The authors are grateful to all the players and students for their participation in this study. Also, thank The Danish Rheumatism Association and Chiroform for their generous support of this study. The funders had no role in the conduct of the study, data analysis, or reporting of the results.

ANONYMOUS TITLE PAGE

The inter- and intrarater reliability and agreement for field-based assessment of scapular control, shoulder range of motion, and shoulder isometric strength in elite youth athletes

ABSTRACT

Objectives: To investigate the intra- and interrater reliability and agreement for field-based assessment of scapular control, shoulder range of motion (ROM), and shoulder isometric strength in elite youth athletes.

Design: Test-retest reliability and agreement study.

Setting: Eight blinded raters (two for each assessment) assessed players on field during two testing sessions separated by one week.

Participants: 162 elite youth handball players with or without a history of previous shoulder pain within the preceding six months.

Main Outcome Measures: Kappa (κ) and prevalence-adjusted bias-adjusted kappa (PABAK) coefficients for scapular control reliability, and 95% limits of agreement (LOA) for ROM and strength agreement.

Results: Scapular control demonstrated substantial to almost perfect reliability (κ 0.67 to 0.84, PABAK from 0.68 to 0.88). Mean strength values ranged from 0.9N/kg to 1.6N/kg, and LOAs ranged from -0.7N/kg to 0.8N/kg. Rotational strength revealed additionally systematic bias between and within rater. No or acceptable systematic bias were evident for ROM and abduction strength measures. Mean values and LOAs for ROM ranged between 39.9° to 52.3°, and from -12.6° to 9.9°, respectively.

Conclusions: Scapular control and ROM can be assessed on the field with acceptable reliability. The threshold for reliable measurements of isometric strength using handheld-dynamometers is high.

High Lights:

- Scapular control and ROM can be assessed with acceptable reliability in a field-based setting.

- Risk of injury threshold for ROM differences must exceed 5° to exceed measurement error.
- Using hand-held dynamometer for strength assessments should be used with caution.

Key words: Handball, hand-held dynamometer, inclinometer, scapular dyskinesis

INTRODUCTION

Sports that require repetitive overhead movements places athletes at high risk for shoulder injury[1]. Consequently, identifying risk factors is important for injury prevention in sport[2]. Previous studies have identified several modifiable risk factors for a shoulder injury in overhead athletes; such as limited dominant shoulder internal rotation range of motion (ROM)[3-5], decreased shoulder external rotation and abduction strength[6,7], and lack of scapular control[3]. However, little is known about the reliability of the physical examination procedures used to measure these factors in sport. Reliability is essential to the clinical utility of a shoulder assessment procedure as well as its use in research, because if the assessments have large measurement error, it may be hard to find an association between the shoulder assessment and future development of shoulder injuries in cohort studies, or to measure the effect of treatment or training over time.

Previous studies[8-13] investigating the inter- and intrareliability of shoulder test outcomes in overhead athletes have reported conflicting findings, with intraclass correlation coefficients (ICC) ranging from 0.13[8] to 0.99[11,12] and kappa (κ) coefficients ranging from 0.08[13] to 0.61[10]. While methodological issues likely account for some variation in reliability outcomes[14,15], study limitations constrain the interpretation of these results. Common shortcomings include the investigation of small cohorts, and inadequate statistical analyses (not supporting ICC reliability coefficients with agreements estimates). Moreover, no other studies have investigated the reliability of both ROM, scapular control, rotational and abduction strength in the same athlete study population. Finally, with only one exception [9], all previous studies have, to the best of our knowledge, been performed in controlled environments i.e. separate rooms which are contrary to the common field-based assessment of athletes.

Therefore, the purpose of the present study was to investigate the intra- and interrater reliability of field-based shoulder testing in a sample of elite youth handball players.

METHODS

Participants

We recruited elite handball players aged 14 to 18 years from five colleges specialized in handball. We excluded players if they reported a history of (a) previous shoulder surgery, (b) previous glenohumeral dislocation, or (c) glenoid labrum tear, (d) rotator cuff tear, or (e) fracture in the shoulder region within the past 6 months, as well as (f) those who experienced pain during assessment procedures.

All participants provided a signed informed consent before study enrolment.

Raters and pilot studies

We recruited eight final year physiotherapy students (two for each shoulder assessment) with no prior experience in performing the assessment of participants. All raters underwent approximately two hours of training in the testing procedures as part of two pilot studies. The first pilot study was performed using twenty physiotherapy students as participants. In this study, we evaluated and refined the test protocols and the feasibility of the procedures based on qualitative feedback from participants as well as an experienced research physiotherapist. In a second pilot study with forty-five handball players aged 14 to 18 years, the raters further refined and evaluated the procedure and the interpretation of test results.

Procedures

All participants attended two testing sessions separated by one week. We performed the testing procedures in rooms or corners available in the participant's own fields before or during a training practice. The same room or corner was used for each testing session, and the second test session was performed at the exact same day of the week and time of the day as the first test session. All testing procedures had to be finished within 1.5 – 2 hours which gave each rater a short time-frame to perform each assessment. Furthermore, the raters performed the assessments with other players around as a disturbing factor.

During data collection, participants rotated between four different test stations in random order. Each station involved assessments of either (1) scapular control, (2) shoulder internal and external rotational isometric strength (3) shoulder internal and external rotational ROM or (4) isometric abduction strength. Scapular control assessment was only evaluated at the first testing session. Two repetitions of each ROM and strength test were performed at each testing session. A practice repetition was first performed to familiarize the participant with the measurement procedure. Raters were blinded to the other raters' results and participants' arm dominance. Additional details of the testing procedures are included as part of online Supplementary Appendix 1.

As a part of the first test session, all participants answered a baseline questionnaire that established their arm dominance and weight.

Scapular control

The participants performed five repetitions of full shoulder flexion and abduction while holding either a 3 kg (for male) or 2 kg (for female) hand weight. Girls were only allowed to wear normal bra, and not sports bras. Scapular dyskinesis was defined as the presence of either winging or dysrhythmia as described by McClure et al. [13]

Each shoulder was evaluated independently and classified as (a) normal scapular control (the scapula is stable with minimal motion during the initial 30 degree to 60 degree humerothoracic elevation, then smoothly and continuously rotates upward during elevation and smoothly and continuously rotates downward during humeral lowering. No evidence of winging is present [13]), (b) subtle scapular dyskinesis (mild or questionable evidence of either dysrhythmia or winging, not consistently present [13]) or (c) obvious dyskinesis (striking, clearly apparent abnormality like dysrhythmias or winging of 2.54 or greater displacement of scapular from thorax, evident on at least 3/5 trials[13]) during shoulder flexion and abduction movements. The final evaluation of scapular control was based on combined flexion and abduction test movements as described by McClure et al.[13]. If both motions were rated as normal or subtle, the final rating was normal. If both motions were rated as subtle dyskinesis, the final rating was subtle dyskinesis, and if either motion was rated as obvious dyskinesis, the rating was obvious dyskinesis.

Internal and external range of motion

The ROM test protocol was based on a procedure described previously[3]. We measured shoulder internal and external ROM using a single digital inclinometer (Pro 3600 Digital Protactor, Level developments) with the participant supine on a portable table and the shoulder abducted to 90°, and the elbow flexed to 90°. A ruler was taped to the inclinometer to ensure correct placement at the midpoint between the ulnar styloid and olecranon. A folded towel was placed under the distal humerus to ensure alignment of the upper arm in the frontal plane. The rater palpated the coracoid process of the involved scapula and rotated the participant's shoulder to end range. End range internal and external ROM was defined as the point at which the coracoid process was felt to move in either anterior direction (internal rotation) or posterior/superior direction (external rotation). The rater stabilized the shoulder in this position by placing their medial forearm (the arm closest to the examination table) on the participant's upper arm and exerting downward pressure while gripping the participant's forearm [3] (Appendix 1).

Isometric internal and external strength

Maximum isometric internal and external rotational strength was assessed using a handheld dynamometer (Commander™ Muscle Tester, JTECHmedical) and from a modified version of the protocol reported by Hurd et al.[16]. Participants were positioned supine with their shoulder abducted to 90° and in neutral rotation, and elbow flexed to 90°. A strap was placed across the participant's anterior pelvis along the anterior superior iliac spines to fixate the lower trunk to the table.

The raters used both arms, which were straight (no flexion at the elbow) to avoid counter pressure, and placed their folded hands around the anterior (internal rotation) or posterior (external rotation) part of the wrist. The dynamometer was strapped around the fingers on the medial hand so that they could use the lateral hand to stabilize the equipment at its correct placement on the wrist as shown in Appendix 1. The participant was instructed to maintain a maximal contraction against the dynamometer for 5 seconds.

Abduction strength

Shoulder maximum isometric abduction strength was performed with the participant in the 'full-can' supraspinatus test position as described by Reinold et al.[17]. A 30-degree angle was marked on the floor with tape to align the participant's shoulder in the plane of the scapula. Their shoulder was positioned in 90^{o} of abduction using a goniometer, with the thumb pointing upwards and with the arms in the scapular plane position. The rater was seated in a chair with arms elevated and extended at the elbow. A handheld dynamometer (Commander™ Muscle Tester, JTECHmedical) was positioned 1 cm proximal from the line of the radiocarpal joint (Appendix 1). The participant was instructed to exert a maximal contraction against the dynamometer and rater while maintaining a maximum contraction for 5 seconds.

Statistical analysis

All statistical analyses were conducted in Stata version 14.1 software (StataCorp, College Station, TX, USA). We calculated means and standard deviations across participants and raters for all dependent continuous variables based on the first test session, which was applied as normative reference values. Isometric strength data is presented in newton (N), and N normalized to body weight (N/kg). In addition to the absolute ROM measures, we calculated the difference between the dominant and non-dominant arm.

We estimated the intertester reliability of scapula control with Cohen's kappa coefficients (κ). To assist the interpretation of κ outcomes, we also calculated indices of prevalence and bias and prevalence-adjusted bias-adjusted kappa coefficients (PABAK)[18]. Benchmarks suggested by Landis and Koch[19] were used to interpret the κ and PABAK outcomes (>0.81, almost perfect; 0.61 to 0.80, substantial; 0.41 to 0.60, moderate; 0.21 to 0.40, fair; 0.00 to 0.20, slight; and <0.00, poor).

We calculated the reliability and agreement calculations for continuous measures in two ways. Preferably, we wanted to use a similar approach as recommended by Hayen et al.[20], as this allows us to assess inter- and intrarater reliability simultaneously, and take the repeated measures of our design into account. In this approach, we analyzed differences using a mixed two-way analyses of variance (ANOVA) for repeated measurements. Rater and period were entered as fixed effects, and the following were entered as random effects: participant x rater, participant x period and residuals. Based on the variance components, we calculated Bland and Altman bias and 95% limits of agreement (LOA) statistics and ICCs for the following comparisons: 1) within the same rater and day, 2) within the same rater on different days, and 3) within the same day and between different raters. Formulas can be found in supplementary material as Appendix 2. However, the applied statistical model assumes no systematic differences between the rater's two repeated measurements within a day, which was not the case for all the strength measures in our study. For these assessments, we, therefore, calculated LOA[21,22] and ICC based on the mean between each rater's two repeated measurements for the inter- and intrareliability between days. ICCs were calculated using a two-way mixed absolute agreement model (ICC 3,1)[23]. Benchmarks suggested by Fleis were used to interpret ICC outcomes: (a) >0.90 = excellent reliability; (b) 0.80 to 0.89 = good; (c)

0.70 to 0.79 = moderate; and finally (d) <0.70 = low reliability [24]. LOAs were interpreted as the minimal detectable change (MDC)[25].

The number of participants included in our analysis was based on the formula for limits of agreement described by Bland & Altman [22]: N= (2*1.96*s/w)^2 , where s is the standard deviation and w the width of the LOAs. We only applied this calculation for the ROM procedures. SD was set to be 11 based on a study of TROM on badminton players[26]. Since the inclinometers are very sensitive we set our acceptable LOA to 5. Based on these assumptions, we therefore required 74 players for each gender.

RESULTS

We enrolled 162 participants (82 girls) in the study, and the number of participants included in the analyses for each assessment for the dominant arm is listed in Figure 1. The demographic characteristics of the sample are presented in Table 1. Normative reference values for ROM and strength measurements are presented in the Supplementary material as Appendix 3.

(Please place Figure 1 and Table 1 around here)

Scapular control

The intertester reliabilities of scapular control assessments are presented in Table 2. The raters' individual results for the final combined rating are found in Supplementary material as Appendix 4. The results demonstrated substantial κ coefficients (range 0.59 to 0.96) except for abduction movements for the non-dominant arm, which revealed moderate κ coefficients of 0.47 to 0.49. PABAK agreement ranged between 0.68 and 0.80. Based on both raters' assessments, the prevalence proportion of obvious scapular dyskinesis in the dominant arm was 10% for girls and 39% for boys. For the non-dominant arm, the prevalence proportion was 21% for the girls and 39% for the boys (Appendix 4). We estimated the indices of prevalence and bias[18] for the girls dominant arm to -0.80 and -0.02, respectively, and for the boys dominant arm to -0.21 and 3.91, respectively. For the non-dominant arm, we estimated the indices of prevalence and bias for the girls to -0.57 and 0.04, and for the boys to -0.19 and 0 (Appendix 4).

(Please place Table 2 around here)

Range of motion

Reliability and agreement results for ROM assessments are summarized in Table 3. Rater 1 systematically measured some degrees lower than Rater 2 in 6 out of 8 (6/8) measurements. Intrareliability between days revealed systematic bias of approximately 1° in 4/8 measurements. For both internal and external range of motion, we found the narrowest LOAs within the same rater compared to between rater measurements, in which external rotation revealed the widest LOAs. The systematic differences between raters decreased when looking at the difference between dominant and non-dominant calculations compared to the absolute values, but the LOAs were approximately the same except for the intrarater measurements within day, which revealed wider LOAs. Mean values ranged between 39.9° and 52.3° (Appendix 3).

(Please place Table 3 around here)

Isometric rotational strength

Rater 3 was the only rater who in 9/16 assessments had between-day bias. Between rater assessments demonstrated systematic bias in all male assessments. Intra- and interrater LOAs were approximately the same ranging from -0.5 N/kg to 0.9 N/kg (Table 4). Mean values ranged between 1.3 N/kg and 1.6 N/kg (Appendix 3).

Abduction strength

One significant difference was found in the intrarater assessments and none between raters. LOAs ranged from -0.4 N/Kg to 0.4 N/Kg (Table 4) and mean values from 0.9 N/kg to 1.2 N/kg (Appendix 3).

(Please place Table 4 around here)

DISCUSSION

The present study is the first to establish the reliability and agreement of scapular control, and shoulder ROM and strength assessments in the same large study population in field-based conditions.

Scapular control

We identified greater κ values (0.67 to 0.84) than those presented in previous studies (κ = 0.55 to 0.58) [13]. In our study, all boys performed the assessments with a 3-kg dumbbell and all the girls with a 2-kg dumbbell, which is slightly higher weights than described by McClure et al. [13]. In addition, McClure et al. based the choice of dumbbell on body weight and not on gender, which may have influenced the comparison of the two study results. There were slight differences between the κ and PABAK values, which indicate that the κ values were influenced by either bias or the prevalence of obvious dyskinesis. The prevalence of obvious scapular dyskinesis in the dominant arm for the girls was only 10%. This number is much lower in what we found in our cohort study [27] using the same procedure as described in this study. Here the prevalence proportion for the dominant arm in girls was 30% (results not published). It is likely that the low prevalence proportion has influenced our results for the girls as the prevalence influence the expected agreement by chance.

It has been argued that scapular control should be dichotomized (e.g., absent or present) rather than categorized (normal control, slight dyskinesis, obvious dyskinesis)[28]. Our findings demonstrate that dichotomization increases interrater agreement when compared to three-option categorization. Consequently, dichotomization of scapular control into normal (normal+subtle dyskinesis) or obvious dyskinesis may be more suitable for research and applicable for clinical use.

Range of motion

No clinically relevant systematic error (bias) was identified for intrarater agreements (the difference was 1° or below). Intrarater LOAs between days ranged from -8.4° to 9.9°, which means that almost 10°s change is required to be 95% certain that the change between the measurements is not due to variability of measurement error if the same rater repeats the measurement. Interrater agreements revealed systematic error (bias) of between 3.6° and 6.7°s, and slightly higher LOAs for particularly external rotation.

Injury risk factor studies have included differences in shoulder ROM between dominant and non-dominant arms as a potential predictor of injury [3,4]. Our results demonstrate that using this calculation reduced the amount of systematic error in the inter-rater assessments; however, based on the LOA, a 10°s change is still required to be sure a change in the measurement is not due to measurement error. In handball, 5°s change in Total ROM has been reported to be associated with reduced odds for shoulder injury [Odds Ratio (OR) :0.77 (95% CI 0.56 to 0.995)][3]. Thus, it can be argued that 5°s represent a clinical important change. Unfortunately, the study by Clarsen et al. only reports ICCs and not LOAs as reliability measures. However, according to our study, a difference of 5°s may be a low threshold for a clinical change as there is a reasonable chance that this is due to measurement error.

In previous reliability studies of passive ROM in overhead athletes, only one study undertook analyses beyond the calculation of ICCs. Boon et al.[8] reported MDC values ranging from 18.23° to 27.55° for intrarater reliability and 22.14° to 25.21° for inter-rater reliability, which are considerably higher than our LOA estimates. However, they also reported greater absolute rotational maximum values than ours, which affects the MDC[24]. These dissimilarities may be explained by differences in measurement technique and instruments as Boon et al. used a goniometer to measure shoulder ROM, while we used a digital inclinometer[24].

Rotational strength

We identified significant systematic errors in 10/16 assessments in the interrater analysis, and in 9/16 measurements in the intrarater analysis for rater 3, whereas rater 4 did not demonstrate any significant systematic errors.

Limits of agreements for the rotational strength measures without systematic differences demonstrated an individual range in both internal and external rotation (-48 N to 51 N). This is approximately 42% – 56% of the estimated mean reference values. In addition, it is more than 5 times higher than 10 N, which can be argued to represent the minimal clinically important difference, as Clarsen et al. have reported a 29% reduced odds [OR:0.71 (95% CI 0.44 to 0.99)] for substantial shoulder problem per 10 N increase in external rotational strength[3]. Clarsen et al. performed this test differently with the arm at 0 °s of abduction [3], which might have an influence on the clinical importance value, but still given the high

LOAs it is extremely difficult to estimate an actual clinical difference using this approach in youth handball.

Fieseler et al.[11] is to our knowledge the only other study which has investigated the isometric intrarater reliability of internal and external rotation using HHD between 7 days. They report LOAs ranging from -17.0 N to 19.4 N for internal rotation and -18 N to 15 N for external rotation in the throwing arm, which are narrower LOAs than the LOAs reported in our study, but still almost twice as high as what might be the clinical relevant difference[11]. A possible explanation of these dissimilarities may be that both of our raters were female, who had trouble holding the position when testing some of the strong males, as demonstrated by the systematic bias for primary male assessments. A previous study has argued that when the strength of the muscle group being tested exceeds the capacity of the assessor to hold against or stabilize the assessed person, the force measured will represent the limitations of the rater and not the strength of the assessed subject [9]. Hand held dynamometry is easy to apply, but the test procedures have to be improved so that the measurements do not rely on the strength of the rater. A possible solution might be to attach the HHD to a suction cup [29]. We, therefore, modified our procedures to include external belt-fixation, and re-evaluated it in a small sample of 17 male u-18 handball players [27]. This approach narrowed the LOA by up 50% compared to the intrareliability results presented in this paper, which indicate, that this could be an applicable and reliable approach to use for field-based assessments of rotational strength. However, it has to be further investigated in a larger sample.

Abduction strength

No significant differences existed between the two raters and within raters in their measurement of abduction strength. Still, the MDC were around 0.4 N/kg in the worst cases, which is high according to mean values ranging between 0.9 N/kg to 1.2 N/kg.

Normalizing the N values to weight changed some of the reliability results slightly in both the rotational and abduction strength measurements. We have presented both measures to be able to compare N outcomes with previous studies and to provide the reliability results for the normative values normalized to weight (Appendix 3) which is the advised strength measure to use in risk factor studies of overhead athletes[30].

The only other study investigating the reliability of abduction strength is also the only study besides ours assessing reliability measures on-field[9]. Unfortunately, they only report ICC, which makes comparisons difficult. One could speculate that the reason for the wide LOAs in both abduction and rotational strength is the result of factors within the player such as the players' motivation or fatigue, as other agreement studies from other populations also demonstrate relatively large MDC using a HHD[24,31]. In our study, the players rotated between several test stations and it is possible that this influenced the players' motivation and fatigue. However, a previous study investigating the reliability of abduction strength conclude that a strength gain is more reliable to use than a HHD because it has the advantage of having the participant to pull up against a plate that is stabilized by his body weight instead of relying on the strength of the assessor[9]. Thus, eliminating the random errors from the raters by attaching the HHD to a suction cup instead of using the rater’s arms [29] might improve our results. As for the rotational strength procedures, we, therefore, modified our abduction strength procedures to include external belt-fixation, and re-evaluated it in the same sample as described above [27]. Again, this approach narrowed the LOA by up 50% compared to our intrareliability results presented in this paper, and was much easier to perform for the physiotherapist. The reliability of this procedure has to be further established due to the small sample size in that study[27].

Statistical approach

Unfortunately, very few reproducibility studies in overhead sport have reported other reliability values than ICC. It has been argued[25,32] that ICC is a poor estimate to solely conclude on because high ICC does not necessarily mean that a test is reliable as well as low ICC does not necessarily mean that a test is unreliable[13]. In situations with a homogeneous sample in which there is little variability among subjects' scores, it is difficult to obtain a high ICC despite low measurement error, whereas high ICC may be reported in a more heterogeneous sample with greater variation between subjects' scores in which the ICC will be scarcely affected by measurement error[25]. Absolute reliability measures as SEM, LOA or MDC are much easier to interpret and more applicable for the clinicians in the field as they reflect the agreement and error in the relevant value of the measurement. Our results are a good example of how conclusions based solely on ICCs can be misleading (Table 3 & 4). Intraclass correlation coefficients for the strength measurements demonstrated substantial and almost perfect reliability but the poorest agreements, whereas the ROM measurements revealed poor reliability based on the ICCs but more acceptable LOAs and no or small systematic difference for particularly intrarater measures.

Another drawback to the use of ICC and comparison between studies is that of the mentioned seven reliability studies in overhead sport[3,8,9,11,12,33,34], only two studies[3,12] have stated the statistical ICC model they have applied. In our study, we aimed to include all measurements of each rater in the ANOVA analysis[20], which we believe gives a more accurate picture of the reliability and measurement errors. However, the current model assumes no systematic differences between the rater's two repeated measurements within the day, which was not fulfilled for all the strength measures in our study. We, therefore, had to use the mean of the measurements[23], which improves reliability estimates compared to those derived from single measures[25]. In future studies, the applied ANOVA statistical model can be extended to accommodate for within day systematic differences.

Limitations

In a clinical setting, or in studies seeking to establish modifiable non-participating risk factors for shoulder injuries, it is often only possible to test players in a short time-frame before or during training sessions in rooms or corners available in that particular field, and with players around as a disturbing factor. The primary study strength was the pragmatic measurement approach that reflected the real-world application, thus enhancing external validity of the results. However, due to the short time frame for testing, the observers did not manage to test all players in all the tests. For abduction strength and scapula control, rater 5, 6 and 8 were absent from one test day. The smaller study sample for these particular assessments may have influenced the results.

Furthermore, these results are based on only two novice raters. We recruited novice physiotherapists mainly due to practical and economic reasons, however, this also reflects the "real-world" scenario in youth handball as very few youth handball teams have an experienced physical therapist connected. Nevertheless, this limits the generalizability of our results. Larger studies including several raters, among them, more experienced raters should be conducted before conclusive clinical recommendations can be made.

Clinical relevance

Scapular control, range of motion and isometric strength measures have all been used to identify risk factors for shoulder injuries in sport [3-7], and are used in clinical practice to measure the effect of treatment or training over time. Our results highlight the importance of taking the measurement error for continuous measures into account when interpreting results in risk factor studies and clinical practice. Such measurement errors may explain why it has been difficult to define a cut point in which continuous variables are translated into a dichotomous risk factor that can distinguish whether a player is at increased risk or not [35]. It should further be emphasized that clinicians or raters need to be trained before using these tests in practice, and it is recommended that clinicians and raters routinely perform intra- and interrater agreement tests to reduce the measurement errors.

CONCLUSIONS

Scapular control can be assessed in elite youth athletes with acceptable reliability in a field-based setting. Shoulder range of motion can be assessed with acceptable intrareliability within day. However, intrareliability between days and interreliability demonstrated greater levels of measurement error. This emphasizes that the risk of injury threshold for ROM differences used in risk factor studies must exceed the commonly use of 5 degrees to ensure observed changes are not due to measurement error.

Using hand-held dynamometer for isometric shoulder rotation and abduction strength assessments should be used with caution due to high threshold for reliable measures, and future studies should investigate new procedures to measure shoulder isometric strength measures in athletes.

REFERENCES

1. Wilk KE, Obma P, Simpson CD, Cain EL, Dugas JR, Andrews JR. Shoulder injuries in the overhead athlete. *J Orthop Sports Phys Ther.* 2009;39(2):38-54.
2. Finch C. A new framework for research leading to sports injury prevention. *J Sci Med Sport.* 2006;9(1-2):3-9; discussion 10.
3. Clarsen B, Bahr R, Andersson SH, Munk R, Myklebust G. Reduced glenohumeral rotation, external rotation weakness and scapular dyskinesis are risk factors for shoulder injuries among elite male handball players: a prospective cohort study. *Br J Sports Med.* 2014;48(17):1327-1333.
4. Shanley E, Kissenberth MJ, Thigpen CA, et al. Preseason shoulder range of motion screening as a predictor of injury among youth and adolescent baseball pitchers. *J Shoulder Elbow Surg.* 2015;24(7):1005-1013.
5. Wilk KE, Macrina LC, Fleisig GS, et al. Correlation of glenohumeral internal rotation deficit and total rotational motion to shoulder injuries in professional baseball pitchers. *Am J Sports Med.* 2011;39(2):329-335.
6. Byram IR, Bushnell BD, Dugger K, Charron K, Harrell FE, Jr., Noonan TJ. Preseason shoulder strength measurements in professional baseball pitchers: identifying players at risk for injury. *Am J Sports Med.* 2010;38(7):1375-1382.
7. Edouard P, Degache F, Oullion R, Plessis JY, Gleizes-Cervera S, Calmels P. Shoulder strength imbalances as injury risk in handball. *Int J Sports Med.* 2013;34(7):654-660.
8. Boon AJ, Smith J. Manual scapular stabilization: its effect on shoulder rotational range of motion. *Arch Phys Med Rehabil.* 2000;81(7):978-983.
9. Burnham RS, Bell G, Olenik L, Reid DC. Shoulder abduction strength measurement in football players: reliability and validity of two field tests. *Clin J Sport Med.* 1995;5(2):90-94.

10. Ellenbecker TS, Kibler WB, Bailie DS, Caplinger R, Davies GJ, Riemann BL. Reliability of scapular classification in examination of professional baseball players. *Clin Orthop Relat Res.* 2012;470(6):1540-1544.

11. Fieseler G, Molitor T, Irlenbusch L, et al. Intrarater reliability of goniometry and hand-held dynamometry for shoulder and elbow examinations in female team handball athletes and asymptomatic volunteers. *Arch Orthop Trauma Surg.* 2015;135(12):1719-1726.

12. Kevern MA, Beecher M, Rao S. Reliability of measurement of glenohumeral internal rotation, external rotation, and total arc of motion in 3 test positions. *J Athl Train.* 2014;49(5):640-646.

13. McClure P, Tate AR, Kareha S, Irwin D, Zlupko E. A clinical method for identifying scapular dyskinesis, part 1: reliability. *J Athl Train.* 2009;44(2):160-164.

14. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *PM R.* 2011;3(5):472-479.

15. 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. *Journal of physiotherapy.* 2010;56(1):7-17.

16. Hurd WJ, Kaplan KM, ElAttrache NS, Jobe FW, Morrey BF, Kaufman KR. A profile of glenohumeral internal and external rotation motion in the uninjured high school baseball pitcher, part II: strength. *J Athl Train.* 2011;46(3):289-295.

17. Reinold MM, Macrina LC, Wilk KE, et al. Electromyographic analysis of the supraspinatus and deltoid muscles during 3 common rehabilitation exercises. *J Athl Train.* 2007;42(4):464-469.

18. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46(5):423-429.

19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

20. Hayen A, Dennis RJ, Finch CF. Determining the intra- and inter-observer reliability of screening tools used in sports injury research. *J Sci Med Sport.* 2007;10(4):201-210.

21. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet.* 1995;346(8982):1085-1087.

22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.

23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin.* 1979;86(2):420-428.

24. Cools AM, De Wilde L, Van Tongel A, Ceyssens C, Ryckewaert R, Cambier DC. Measuring shoulder external and internal rotation strength and range of motion: comprehensive intra-rater and inter-rater reliability study of several testing protocols. *J Shoulder Elbow Surg.* 2014;23(10):1454-1461.

25. de Vet HTC, Mokking LB, Knol DL. *Measurement in Medicine. A pratical guide.* United states of America: Cambridge University Press, New York; 2011.

26. Couppe C, Thorborg K, Hansen M, et al. Shoulder rotational profiles in young healthy elite female and male badminton players. *Scand J Med Sci Sports.* 2014;24(1):122-128.

27. Moller M, Nielsen RO, Attermann J, et al. Handball load and shoulder injury rate: a 31-week cohort study of 679 elite youth handball players. *Br J Sports Med.* 2017;51(4):231-237.

28. Struyf F, Nijs J, Mottram S, Roussel NA, Cools AM, Meeusen R. Clinical assessment of the scapula: a review of the literature. *Br J Sports Med.* 2014;48(11):883-890.

29. Thorborg K, Bandholm T, Holmich P. Hip- and knee-strength assessments using a hand-held dynamometer with external belt-fixation are inter-tester reliable. *Knee Surg Sports Traumatol Arthrosc.* 2013;21(3):550-555.

30. Cools AM, Vanderstukken F, Vereecken F, et al. Eccentric and isometric shoulder rotator cuff strength testing using a hand-held dynamometer: reference values for overhead athletes. *Knee Surg Sports Traumatol Arthrosc.* 2015.

31. Johansson FR, Skillgate E, Lapauw ML, et al. Measuring Eccentric Strength of the Shoulder External Rotators Using a Handheld Dynamometer: Reliability and Validity. *J Athl Train.* 2015;50(7):719-725.

32. Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med.* 1998;26(4):217-238.

33. Awan R, Smith J, Boon AJ. Measuring shoulder internal rotation range of motion: a comparison of 3 techniques. *Arch Phys Med Rehabil.* 2002;83(9):1229-1234.

34. Wilk KE, Reinold MM, Macrina LC, et al. Glenohumeral internal rotation measurements differ depending on stabilization techniques. *Sports Health.* 2009;1(2):131-136.

35. Bahr R. Why screening tests to predict injury do not work-and probably never will...: a critical review. *Br J Sports Med.* 2016;50(13):776-780.

TABLES AND FIGURES

Table 1. Demographics of study population by sex

	Female (n=82)	Male (n=80)
Shoulder injury within the previous 6 months		
Yes n (%)	30 (37)	19 (24)
No n (%)	52 (63)	61 (37)
Age Group		
U16 n (%)	43 (52)	46 (57)
U18 n (%)	39 (48)	34 (33)
Dominant arm		
Rigth arm n (%)	67 (82)	62 (78)
Left arm n (%)	15 (18)	18 (22)
Player position	*	*
Back players n (%)	38 (47)	30 (38)
Wing players n (%)	18 (22)	28 (35)
Line players n (%)	14 (17)	11 (14)
Goal keepers n (%)	11 (14)	10 (13)
Mean age, years (sd)	16.38 (0.97)*	16.22 (1.16)*
Mean height, cm (sd)	1.73 (0.05)	1.83 (0.07)
Mean weight, kg (sd)	67.95 (8.06)	76.28 (11.65)
Mean years handball experience (sd)	9.52 (2.60)*	8.39 (2.46)
Mean hours weekly handball training (sd)	9.73 (2.78)	8.31 (1.97) †

* 1 missing † 2 missing

Table 2. Interreliability of scapular control for elite youth handball players by sex

Test	Sex	Agreement (%)	KAPPA	PABAK
<u>Dominant arm</u>				
Flexion	Female (n=56)	87.50	0.62	0.75
Flexion	Male (n=69)	89.86	0.73	0.80
Abduction	Female (n=56)	89.29	0.60	0.79
Abduction	Male (n=69)	85.29	0.64	0.71
Scapular control categorized	Female (n=56)	84.06	0.70	0.80
Scapular control categorized	Male (n=69)	90.18	0.67	0.68
Scapular control dichotomized	Female l (n=56)	98.21	0.90	0.96
Scapular control dichotomized	Male (n=69)	85.51	0.70	0.71
<u>Non-dominant arm</u>				
Flexion	Female (n=56)	89.29	0.74	0.79
Flexion	Male (n=70)	90.71	0.77	0.81
Abduction	Female (n=56)	85.71	0.49	0.71
Abduction	Male (n=70)	79.71	0.47	0.59
Scapular control categorized	Female (n=56)	93.75	0.84	0.88
Scapular control categorized	Male (n=70)	84.06	0.67	0.76
Scapular control dichotomized	Female (n=56)	96.43	0.89	0.93
Scapular control dichotomized	Male (n=70)	90.00	0.79	0.80

PABAK= prevalence and bias adjusted kappa. Scapular control categorized = scapular control categorized into (a) normal, (b) subtle, or (c) obvious dyskinesis [25]. Scapular control dichotomized = scapular control dichotomized into (a) normal (normal + subtle), or (b) obvious dyskinesis.

Table 3. Within-day and between-day rater reliability of internal and external rotational range of motion for elite youth handball players by sex

Assessments	Sex	Within-day Same Rater		Within-day Between Rater			Between-day Same Rater		
		LOA	ICC	Bias (95% CI)	LOA	ICC	Bias (95% CI)	LOA	ICC
Dominant arm									
External (°)	Female (n=80)	-4.0 to 4.0	0.82	-5.0 (-5.8 to -4.2)	-12.5 to 2.4	0.36*	-0.7 (-1.4 to 0.0)	-7.6 to 6.2	0.46
External (°)	Male (n=80)	-4.3 to 4.3	0.85	-3.6 (-4.6 to -2.7)	-12.2 to 5.0	0.44*	1.1 (0.1 to 2.2)	-7.6 to 9.9	0.42*
Internal (°)	Female (n=80)	-3.3 to 3.3	0.82	-0.4 (-1.0 to 0.3)	-6.3 to 5.6	0.40	-0.9 (-1.5 to -0.3)	-6.8 to 5.0	0.42*
Internal (°)	Male (n=77)	-3.1 to 3.1	0.87	-0.3 (-1.0 to 0.4)	-6.9 to 6.3	0.45	-0.0 (-0.8 to 0.8)	-6.7 to 6.7	0.43
Non-dominant arm									
External (°)	Female (n=80)	-4.1 to 4.1	0.81	-6.7 (-7.5 to -5.9)	-14.3 to 0.8	0.35*	-0.3 (-1.1 to 0.4)	-7.3 to 6.6	0.46
External (°)	Male (n=75)	-4.3 to 4.3	0.84	-4.1 (-5.0 to -3.2)	-12.6 to 4.3	0.40*	1.0 (0.1 to 2.0)	-7.0 to 9.1	0.45*
Internal (°)	Female (n=80)	-3.2 to 3.2	0.88	-2.1 (-2.9 to -1.4)	-8.9 to 4.7	0.47*	-1.2 (-2.0 to -0.4)	-8.4 to 6.0	0.41*
Internal (°)	Male (n=77)	-3.5 to 3.5	0.87	-1.6 (-2.4 to -0.8)	-9.0 to 5.7	0.40*	-0.2 (-1.0 to 0.6)	-7.1 to 6.8	0.46
Difference dominant and non-dominant arm									
External (°)	Female (n=80)	-5.5 to 5.5	0.79	1.6 (0.6 to 2.7)	-8.4 to 11.6	0.49*	-0.3 (-1.2 to 0.6)	-8.8 to 8.3	0.30
External (°)	Male (n=80)	-6.0 to 6.0	0.75	0.6 (-0.4 to 1.6)	-9.0 to 10.2	0.36	-0.1 (-1.1 to 1.0)	-9.5 to 9.4	0.38
Internal (°)	Female (n=78)	-4.7 to 4.7	0.83	1.7 (0.8 to 2.6)	-7.0 to 10.4	0.41*	0.4 (-0.6 to 1.3)	-8.4 to 9.1	0.42
Internal (°)	Male (n=75)	-4.9 to 4.9	0.84	1.2 (0.2 to 2.3)	-8.3 to 10.7	0.44*	0.0 (-1,0 to 1.1)	-9.1 to 9.2	0.40

LOA = 95% Limits of agreement, CI = Confidence interval, ICC = Intra Class Correlation, * Significant rater effect

Table 4. Within-day and between-day reliability of abduction strength and internal and external rotational shoulder strength for elite youth handball players by sex

Test	Sex	Within-day Between Rater			Between-day Same Rater					
		Bias (95% CI)	LOA	ICC	Bias (95% CI)	LOA	ICC	Bias (95% CI)	LOA	ICC
Rotational strenght dominant arm					Rater 3			Rater 4		
Ext *(N/kg)*	Female	0.0 (-0.1 to 0.1)	-0.5 to 0.5	0.68	0.1 (-0.0 to 0.2)	-0.5 to 0.7	0.49	-0.0 (-0.2 to 0.1)	-0.6 to 0.5	0.50
Ext *(N/kg)*	Male	0.1 (0.0 to 0.2)	-0.5 to 0.7	0.71*	0.1 (-0.0 to 0.2)	-0.5 to 0.7	0.66	0.1 (-0.0 to 0.2)	-0.5 to 0.7	0.62
Int *(N/kg)*	Female	0.1 (0.0 to 0.2)	-0.5 to 0.7	0.65*	0.1 (-0.0 to 0.2)	-0.5 to 0.6	0.66	-0.0 (-0.1 to 0.0)	-0.6 to 0.5	0.55
Int *(N/kg)*	Male	0.3 (0.2 to 0.3)	-0.4 to 0.9	0.76*	0.2 (0.1 to 0.3)	-0.4 to 0.8	0.76*	-0.0 (-0.1 to 0.1)	-0.7 to 0.7	0.55
Ext *(N)*	Female	0.9 (-3.9 to 5.7)	-34.4 to 36.2	0.64	5.6 (-2.5 to 13.8)	-35.7 to 46.9	0.42	-1.8 (-9.9 to 6.2)	-38.9 to 35.3	0.49
Ext *(N)*	Male	6.9 (0.1 to 13.7)	-42.3 to 56.1	0.73*	8.2 (-2.8 to 19.2)	-37.3 to 53.7	0.74	7.6 (-2.3 to 17.6)	-36.0 to 51.3	0.70
Int *(N)*	Female	6.9 (2.4 to 11.3)	-31.1 to 44.8	0.65*	5.0 (-1.6 to 11.6)	-32.1 to 42.1	0.66	-3.2 (-9.3 to 2.8)	-37.3 to 30.8	0.52
Int *(N)*	Male	19.6 (13.9 to 25.2)	-27.2 to 66.4	0.78*	13.1 (4.3 to 21.9)	-30.4 to 56.6	0.79*	0.1 (-9.0 to 9.3)	-48.0 to 48.2	0.60
Rotational strenght non- dominant arm										
Ext *(N/kg)*	Female	-0.0 (-0.6 to 0.6)	-0.5 to 0.5	0.75	0.1 (0.0 to 0.2)	-0.3 to 0.5	0.74*	0.0 (-0.1 to 0.1)	-0.4 to 0.4	0.71
Ext *(N/kg)*	Male	0.2 (0.1 to 0.2)	-0.4 to 0.7	0.75*	0.1 (0.0 to 0.2)	-0.3 to 0.5	0.79*	-0.0 (-0.1 to 0.1)	-0.4 to 0.4	0.70
Int *(N/kg)*	Female	0.0 (-0.0 to 0.1)	-0.5 to 0.6	0.70	0.1 (0.0 to 0.2)	-0.4 to 0.6	0.73*	-0.1 (-0.1 to 0.0)	-0.5 to 0.4	0.62
Int *(N/kg)*	Male	0.2 (0.1 to 0.3)	-0.4 to 0.9	0.75*	0.2 (0.0 to 0.3)	-0.5 to 0.8	0.76*	0.0 (-0.1 to 0.1)	-0.6 to 0.6	0.62
Ext *(N)*	Female	-0.2 (-4.0 to 3.6)	-30.6 to 30.3	0.72	6.1 (0.9 to 11.4)	-21.6 to 33.8	0.71*	2.7 (-2.2 to 7.5)	-23.8 to 29.2	0.70
Ext *(N)*	Male	11.4 (6.6 to 16.1)	-24.3 to 47.0	0.81*	7.4 (0.5 to 14.2)	-22.8 to 37.6	0.83*	-3.5 (-11.0 to 4.0)	-36.6 to 29.6	0.82
Int *(N)*	Female	1.8 (-2.3 to 5.8)	-32.6 to 36.2	0.68	5.5 (-0.1 to 11.2)	-25.7 to 36.8	0.70	-3.5 (-8.8 to 1.9)	-34.3 to 27.3	0.60
Int *(N)*	Male	16.2 (10.3 to 22.0)	-33.1 to 65.4	0.77*	11.5 (2.2 to 20.8)	-36.2 to 59.2	0.78*	2.2 (-6.2 to 10.6)	-42.1 to 46.6	0.70
Abduction strenght dominant arm					Rater 5			Rater 6		
(N/kg)	Female	-0.0 (-0.0 to 0.0)	-0.3 to 0.3	0.84	0.0 (-0.0 to 0.1)	-0.3 to 0.3	0.59	0.0 (-0.1 to 0.1)	-0.3 to 0.3	0.78
(N/kg)	Male	-0.0 (-0.0 to 0.0)	-0.4 to 0.3	0.82	0.0 (-0.1 to 0.1)	-0.4 to 0.4	0.67	0.0 (-0.0 to 0.1)	-0.3 to 0.4	0.73
(N)	Female	0.0 (-2.5 to 2.6)	-17.4 to 17.5	0.47	1.9 (-2.6 to 6.4)	-17.8 to 21.7	0.55	0.7 (-3.8 to 5.1)	-18.8 to 20.1	0.80
(N)	Male	-0.3 (-3.7 to 3.1)	-28.1 to 27.5	0.84	1.9 (-5.3 to 9.1)	-31.4 to 35.2	0.68	3.0 (-2.2 to 8.3)	-25.5 to 31.5	0.88
Abduction strenght non-dominant arm										
(N/kg)	Female	0.0 (-0.0 to 0.1)	-0.2 to 0.3	0.84	0.1 (0.0 to 0.1)	-0.3 to 0.4	0.71*	0.0 (-0.1 to 0.1)	-0.3 to 0.4	0.61
(N/kg)	Male	-0.0 (-0.1 to 0.0)	-0.4 to 0.4	0.79	0.1 (0.0 to 0.1)	-0.2 to 0.3	0.61	-0.0 (-0.1 to 0.1)	-0.4 to 0.4	0.72
(N)	Female	1.5 (-0.9 to 3.9)	-15.4 to 18.5	0.77	4.9 (1.4 to 8.4)	-13.1 to 21.8	0.67	-0.2 (-5.0 to 4.6)	-23.5 to 23.1	0.51
(N)	Male	-1.5 (-5.1 to 2.1)	-29.8 to 26.8	0.87	3.7 (-3.6 to 11.0)	-31.0 to 36.3	0.66	-0.2 (-6.3 to 5.8)	-32.8 to 32.3	0.74

LOA = 95% Limits of agreement, CI = Confidence interval, ICC = Intra Class Correlation, Ext = External, Int = Internal, Ext dif = difference between dominant and non-dominant arm in external rotation, Int dif = difference between dominant and non-dominant arm in internal rotation *Significant rater effect.

Total number of players present at testround 1 (n=162)

Range of motion

Pain-free test Rater1*
- Internal Female (n=78)
- Internal Male (n=75)
- External Female (n=77)
- External Male (n=75)

Pain-free test Rater2*
- Internal Female (n=77)
- Internal Male (n=76)
- External Female (n=76)
- External Male (n=76)

Painful test Rater1
- Internal Female (n=1)
- Internal Male (n=1)
- External Female (n=2)
- External Male (n=1)

Painful test Rater2
- Internal Female (n=1)*
- Internal Male (n=1)*
- External Female (n=2)
- External Male (n=1)

Not tested
- Rater1 Female (n=3)*
- Rater1 Male (n=4)*
- Rater2 Female (n=4)
- Rater2 Male (n=3)

Rotational strength

Pain-free test Rater3
- Internal Female (n=73)
- Internal Male (n=74)
- External Female (n=57)
- External Male (n=60)

Pain-free test Rater4
- Internal Female (n=76)
- Internal Male (n=75)
- External Female (n=60)
- External Male (n=61)

Painful test Rater3
- Internal Female (n=1)
- Internal Male (n=3)
- External Female (n=17)
- External Male (n=17)

Painful test Rater4
- Internal Female (n=3)
- Internal Male (n=3)
- External Female (n=19)
- External Male (n=1)

Not tested
- Rater3 Female (n=8)
- Rater4 Male (n=3)
- Rater3 Female (n=3)
- Rater4 Male (n=2)

Abduction strength

Pain-free test Rater5
- Female (n=49)
- Male (n=68)

Pain-free test Rater6
- Female (n=50)
- Male (n=68)

Painful test Rater5
- Female (n=13)
- Male (n=11)

Painful test Rater6
- Female (n=13)
- Male (n=11)

Not tested
- Female (n=19)
- Male (n=1)

Scapula control

Pain-free test Rater7
- Female (n=75)
- Male (n=79)

Pain-free test Rater8
- Female (n=59)
- Male (n=69)

Painful test Rater7
- Female (n=0)
- Male (n=0)

Painful test Rater8
- Female (n=0)
- Male (n=0)

Not tested
- Rater 7 Female (n=7)
- Rater 7 Male (n=1)

Not tested Rater8
- Female (n=23)
- Male (n=11)

Total number of players present at testround 2 (n=137)

Pain-free test Rater1*
- Internal Female (n=39)
- Internal Male (n=33)
- External Female (n=39)
- External Male (n=32)

Pain-free test Rater2*
- Internal Female (n=34)
- Internal Male (n=30)
- External Female (n=34)
- External Male (n=30)

Painful test Rater1
- Internal Female (n=0)
- Internal Male (n=0)
- External Female (n=0)
- External Male (n=1)

Painful test Rater2
- Internal Female (n=0)
- Internal Male (n=1)
- External Female (n=0)
- External Male (n=1)

Pain-free test Rater3
- Internal Female (n=34)
- Internal Male (n=29)
- External Female (n=31)
- External Male (n=26)

Pain-free test Rater4
- Internal Female (n=38)
- Internal Male (n=30)
- External Female (n=27)
- External Male (n=25)

Painful test Rater3
- Internal Female (n=1)
- Internal Male (n=4)
- External Female (n=14)
- External Male (n=7)

Painful test Rater4
- Internal Female (n=0)
- Internal Male (n=1)
- External Female (n=11)
- External Male (n=6)

Pain-free test Rater5
- Female (n=25)
- Male (n=27)

Pain-free test Rater6
- Female (n=26)
- Male (n=33)

Painful test Rater5
- Female (n=2)
- Male (n=3)

Painful test Rater6
- Female (n=3)
- Male (n=1)

Not tested
- (n=3)

Not tested

Analyzed

- Internal Female (n=80)
- Internal Male (n=77)
- Int dif Female (n=79)
- Int dif Male (n=75)
- External Female (n=80)
- External Male (n=80)
- Ext dif Female (n=78)
- Ext dif Male (n=75)

Interrater
- Internal Female (n=73)
- Internal Male (n=69)
- External Female (n=54)
- External Male (n=53)

Rater 3 Intrarater
- Internal Female (n=33)
- Internal Male (n=26)
- External Female (n=27)
- External Male (n=19)

Rater 4 Intrarater
- Internal Female (n=33)
- Internal Male (n=29)
- External Female (n=23)
- External Male (n=21)

Interrater
- Female (n=46)
- Male (n=66)

Rater 5 Intrarater
- Female (n=21)
- Male (n=23)

Rater 6 Intrarater
- Female (n=21)
- Male (n=31)

- Female (n=56)
- Male (n=69)

Figure 1. Population flow showing the number of participants included, tested and analyzed. (Int+Internal=Internal rotation, Ext+External=External rotation. Ext dif = The difference in external rotation between dominant and non-dominant arm. Int dif = The difference in internal rotation between dominant and non-dominant arm

High Lights:

- Scapular control and ROM can be assessed with acceptable reliability in a field-based setting.
- Risk of injury threshold for ROM differences must exceed 5° to exceed measurement error.
- Using hand-held dynamometer for strength assessments should be used with caution.

The Ethics Committee of the Central Denmark Region exempted the study from full ethical review (167/2012) due to the observational methodological study design. The Danish Data Protection Agency (J. nr. 2012 - 41 -1042) approved the study. All participants provided a signed informed consent before study enrolment.