Measurement properties of handheld dynamometry for assessment of shoulder muscle strength

Sørensen, Lotte; Oestergaard, Lisa Gregersen; van Tulder, Maurits; Petersen, Annemette Krintel

published in
Scandinavian Journal of Medicine and Science in Sports
2020

DOI (link to publisher)
10.1111/sms.13805

document version
Publisher's PDF, also known as Version of record

document license
Article 25fa Dutch Copyright Act

citation for published version (APA)
Sørensen, L., Oestergaard, L. G., van Tulder, M., & Petersen, A. K. (2020). Measurement properties of handheld dynamometry for assessment of shoulder muscle strength: A systematic review. *Scandinavian Journal of Medicine and Science in Sports*, *30*(12), 2305-2328. https://doi.org/10.1111/sms.13805

Received: 17 February 2020 | Revised: 15 July 2020 | Accepted: 3 August 2020
DOI: 10.1111/sms.13805

REVIEW ARTICLE

WILEY

Measurement properties of handheld dynamometry for assessment of shoulder muscle strength: A systematic review

Lotte Sørensen[1,2] | **Lisa Gregersen Oestergaard**[1,3] | **Maurits van Tulder**[1,4] | **Annemette Krintel Petersen**[1,2]

[1]Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark
[2]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[3]DEFACTUM, Central Denmark Region, Aarhus, Denmark
[4]Department of Health Sciences, Amsterdam Movement Sciences research institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Correspondence
Lotte Sørensen, Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Palle Juul-Jensens Boulevard 99, DK- 8200 Aarhus N., Denmark.
Email: lotsoere@rm.dk

Like any assessment tool, handheld dynamometry (HHD) must be valid and reliable in order to be meaningful in clinical practice and research. To summarize the evidence of measurement properties of HHD for the assessment of shoulder muscle strength. Cochrane Central Register of Controlled Trials (CENTRAL), PubMed, EMBASE, and PEDro were searched up to February 2020. Inclusion criteria were studies (a) evaluating HHD used on the glenohumeral joint, (b) evaluating measurement properties, and (c) included individuals ≥ 18 years old with or without shoulder symptoms. Exclusion criteria were studies (a) including patients with neurologic, neuromuscular, systemic diseases, or critical illness or bed-side patients and (b) that did not report the results separately for each movement. In total, 28 studies with 963 participants were included. The reliability results showed that 98% of the intraclass correlation coefficient (ICC) values were ≥0.70. The measurement error showed that the minimal detectable change in percent varied from 0% to 51.0%. The quality of evidence was high or moderate for the majority of movements and type of reliability examined. Based on the evidence of low or very low quality of evidence, the convergent validity and discriminative validity of HHD were either sufficient, indeterminate, or insufficient. The reliability of HHD was overall sufficient, and HHD can be used to distinguish between individuals on the group level. The measurement error was not sufficient, and evaluation of treatment effect on the individual level should be interpreted with caution.

KEYWORDS
handheld dynamometry, measurement error, reliability, shoulder, strength, validity

1 | INTRODUCTION

Muscle strength testing is an important consideration in both clinical practice and research and is frequently used in both athletes and patients with shoulder disorders.[1-3] Muscle strength evaluation can assist clinicians to quantify the degree of impairment and guide treatment.[1,4] Thus, assessments of muscle strength are commonly performed before and after an intervention to quantify treatment efficacy.[3]

Manual muscle testing, as a test method for assessing muscle strength, has demonstrated poor reliability compared with more objective methods such as handheld dynamometry

PROSPERO registration number: CRD42017054027.

Protocol publication: DOI:10.1080/10833196.2017.1392673.

(HHD) and isokinetic dynamometry (ID).[5-7] Although ID is often considered the criterion standard in muscle strength assessment, HHD has the advantage to be portable, inexpensive, easy to use, and minimally time-consuming.[1,4]

Prior to monitoring muscle strength in a clinical setting or in research, the measurement properties need to be established.[8] Relevant measurement properties for an objective instrument like HHD are validity, reliability, measurement error, and responsiveness.[9-11] The validity can be examined by comparing the instrument to another similar instrument (convergent validity) or by comparing groups who are expected to be different (discriminative validity).[9] Reliability (relative reliability) and measurement error (absolute reliability) reflect the consistency of a measurement method.[12] Reliability determines whether subjects can be distinguished from each other despite measurement errors, and is highly dependent on the heterogeneity of the study population.[8,12,13] Measurement error assesses how close to each other the results of the repeated measurements are. It quantifies the systematic and random error of a score that is not attributed to true changes in the construct to be measured.[13,14] Responsiveness refers to the ability of an instrument to detect change over time.[9]

Few systematic reviews have evaluated the measurement properties of HHD for muscle strength testing of the shoulder. Schrama et al[1] summarized intra-examiner reliability of HHD in all joints of the upper extremity and found conflicting results or unacceptable reliability for the assessment of shoulder strength. In this review, measurement error was not addressed and both children, adults, and patients with neurologic or systemic diseases were included resulting in a large heterogeneous population. Children vs adults and different disease groups could potentially show a different reliability. Thus, a more homogeneous population is desired. Stark et al[4] evaluated the convergent validity by comparing HHD and ID and found Pearson's correlation coefficients from 0.28 to 0.88 for the shoulder muscles. However, the literature search is from 2010 and an update seems relevant.

Therefore, a systematic review of all relevant measurement properties of HHD, including measurement error, is useful to athletes, musculoskeletal researchers, and clinicians. The aim of this systematic review was to summarize the evidence of measurement properties of HHD for the assessment of shoulder muscle strength of the glenohumeral joint in non-neurological participants.

2 | METHODS

This systematic review was conducted according to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology for systematic reviews of patient-reported outcome measures (PROMs),[15] and, when needed, adapted to the purpose of evaluating the quality of objective measurement instruments. This guideline is based on existing guidelines for reviews, such as the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement (PRISMA),[16] the Cochrane Handbook for systematic reviews of interventions,[17] and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) principles.[18]

2.1 | Deviations from the protocol

Before starting the review, a protocol was written and registered in the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42017054027) and published https://doi.org/10.1080/10833196.2017.1392673. This protocol reflects the original purpose of the review, which was to summarize the evidence of both HHD and ID. As the two types of dynamometers are not sufficiently similar to be combined, the results were intended to be reported separately but in one review. During the process, we considered the data to be too comprehensive to be reported in a single review. Therefore, we decided to divide the review into two to make it more manageable: one review focusing on HHD and another on ID. The focus of the current review is on HHD.

2.1.1 | Search strategy

The following electronic databases were searched from their inception to February 2020: Cochrane Central Register of Controlled Trials (CENTRAL), PubMed, EMBASE, and PEDro. The electronic search strategies were developed with the assistance of an information specialist using MeSH (PubMed), Thesaurus (EMBASE), and free-text words. These terms were combined with the validated sensitive methodological search filter to identify studies of measurement properties in PubMed[19] and an adapted version of the filter made for EMBASE. The search strategy for PubMed is presented in Appendix 1. Reference lists and citations of the included studies and relevant reviews were hand-searched for additional studies. No publication period or language restrictions were applied. Only studies published as a full-text original article were included.

2.1.2 | Procedures

Two review authors (AKP and LS) independently conducted the study selection, data extraction, quality assessment, and rated the results. Disagreements were resolved through discussion, and when needed, a third review author (LGO) was consulted.

2.1.3 | Study selection

Inclusion criteria for studies assessed in this review were (a) evaluating HHD used on the glenohumeral joint, (b) evaluating measurement properties, and (c) included individuals ≥18 years with or without shoulder symptoms.

Exclusion criteria were studies that (a) included patients with neurologic, neuromuscular, systemic diseases, or critical illness or bed-side patients and (b) did not report the results separately for each movement.

Full text of all potentially relevant studies was obtained to identify studies meeting the inclusion criteria. The reasons for exclusion of retrieved full-text articles were recorded (data not shown).

2.1.4 | Data extraction

Characteristics of the instrument, included study population (age, gender, healthy/symptomatic individuals), and test procedure, and results of the measurement properties were extracted from the included studies.

2.1.5 | Quality assessment

The methodological quality of each study was assessed using the COSMIN risk of bias checklist.[15,20] The checklist consists of 10 boxes. The measurement properties assessed in each article were evaluated using the corresponding COSMIN box. Each box contains items dealing with design aspects and statistical methods. The methodological quality was rated as either "very good," "adequate," "doubtful," or "inadequate." To determine the overall quality of each study on a measurement property, the lowest rating in the box was used ("the worst score counts" principle).

The results of each study were rated as either sufficient (+), insufficient (-), or indeterminate (?) based on the described criteria for good measurement properties.[15,20,21]

Construct validity was rated as sufficient if the results were in accordance with the hypothesis defined by the review team.[15] The correlation between compared instruments (convergent validity) had to be ≥0.70 or show no significant differences between instruments.[15] The comparison between groups that were expected to be different (discriminative validity) had to be significantly different.[15]

Reliability was rated as sufficient if intraclass correlation coefficient (ICC) was ≥0.70. Measurement error was rated as sufficient if minimal detectable change in percent (%MDC) was smaller than the minimal important change (MIC).[15] Some studies indicated a change of 10% to 15% as being clinically relevant.[3,22] Based on available literature, we set an average MIC to 15%. Furthermore, we conducted a sensitivity analysis by setting MIC to 10% and 20%, respectively.

In studies using $MDC_{90} = SEM*1.65*\sqrt{2}$,[23,24] it was converted to $MDC_{95} = SEM*1.96*\sqrt{2}$[12] before rating %MDC.

For studies only reporting MDC as absolute values expressed in Newton (N), the %MDC was calculated by dividing MDC with the mean of test and retest, if the study presented sufficient data.

2.1.6 | Evidence synthesis

When results from different studies or across reported test conditions (days 1 and 2, testers 1 and 2, dominant and non-dominant arms, affected and unaffected arms, throwing and non-throwing arms, position of the body, position of the arm, highest and average values, within and between sessions, healthy subjects and patients, different ICC models) were consistent, these results were summarized to determine the overall evidence of the measurement properties of HHD.[15] The results for each movement of the shoulder and for intra- and inter-rater assessment were reported separately. This differentiation was made to present the results in an operational way for clinicians, who are generally interested in specific movements or a specific type of reliability.

The ICC and MDC results were reported as a range between minimal and maximal values and as the proportion of reported ICC estimates ≥0.70 or %MDC estimates ≤15%.

The summarized results of all included studies were rated against the criteria for good measurement properties to determine whether the overall measurement property was sufficient (+), insufficient (-), inconsistent (±), or indeterminate (?). The ratings as either sufficient or insufficient were based on the principle that at least 75% of the results should meet the criteria.[9] The results were rated as "sufficient" if ≥75% of the ICC or %MDC values met the criteria, as "insufficient" if ≤25% of the ICC or %MDC values met the criteria, and as "indeterminate" if between 25% and 75% of the ICC or %MDC values met the criteria.

The sensitivity analysis reported the proportion of reported %MDC estimates ≤10% and %MDC estimates ≤20%.

After summarizing the results, the quality of evidence was rated according to the GRADE approach, modified for reviews of measurement properties.[9] The quality of evidence was graded as "high," "moderate," "low," or "very low."[15,25] The grading was based on risk of bias, inconsistency, imprecision, and indirectness.[15] Risk of bias refers to the methodological quality of the studies; inconsistency to the unexplained inconsistency of results across studies; imprecision to the total sample size of the included studies; and indirectness to the circumstance where evidence comes from different populations or context.[15]

3 | RESULTS

The electronic search strategy for both HHD and ID yielded 8054 hits: 1686 duplicates were excluded and 6368 remained. Based on the title and abstract, 6268 studies were excluded, leaving 100 studies for full-text assessment. Another 59 studies were excluded after evaluating full text; four additional studies were identified through reference checking. In total, 28 studies on HHD were identified (Supplementary Figure S1).

Characteristics of the included studies are shown in supplementary Table S1. The study populations included 806 healthy subjects in 25 studies,[22-24,26-47] 145 patients with shoulder symptoms in five studies,[6,32,35,38,48] and 12 patients with chronic obstructive pulmonary disease in one study.[49] Appendixes 2 and 3 show the results of the measurement properties extracted for each study. Rating of the results and risk of bias assessment are presented in Supplementary Table S2.

3.1 | Reliability

Reliability was investigated in 28 studies.[6,22-24,26-49] Table 1 shows that 98% of the reported ICC values were ≥ 0.70, and therefore, the overall rating of the results was "sufficient" for all movements and type of reliability investigated. The quality of evidence was "high" for internal rotation, external rotation, abduction, and flexion and "high" to "very low" for adduction, and extension.

3.2 | Measurement error

Measurement error was investigated in 14 studies.[22-24,26-29,31-33,35,37,40,48] Table 2 shows that the %MDC values ranged from 0% to 51.0% and none of the movements and type of reliability investigated was rated as

TABLE 1 Summary of findings for reliability of handheld dynamometry reported separately for each movement and for each type of reliability

Movement and type of reliability	Number of studies, number of participants	Summary of results: ICC, range	Summary of results: proportion of ICC ≥ 0.70, %	Overall rating of results	Quality of evidence
IR Intra-rater	12 studies, n = 237 [6,22-24,27,31,32,34,38,40,41,47]	0.57-1.00	97	Sufficient	High
IR Inter-rater	9 studies, n = 243 [6,22,23,27,32,34,38,40,47]	0.72-0.99	100	Sufficient	High
ER Intra-rater	16 studies, n = 561 [6,22-24,29,31-34,37,38,40,41,46-48]	0.70-1.00	100	Sufficient	High
ER Inter-rater	11 studies, n = 316 [6,22,23,32-34,37,38,40,47,48]	0.64-0.98	91	Sufficient	High
Abd Intra-rater	12 studies n = 472 [28,29,32,34,37-40,42,44,46,48]	0.55-0.98	95	Sufficient	High
Abd Inter-rater	10 studies n = 340 [28,32,34,37,38,40,42,44,45,48]	0.77-0.98	100	Sufficient	High
Add Intra-rater	2 study n = 41 [32,38]	0.87-0.98	100	Sufficient	Very low
Add Inter-rater	2 study n = 117 [32,38]	0.86-0.98	100	Sufficient	Moderate
Flex Intra-rater	7 studies n = 124 [6,24,28,32,37,40,49]	0.86-0.97	100	Sufficient	High
Flex Inter-rater	6 studies n = 208 [6,28,32,35,37,40]	0.86-0.97	100	Sufficient	High
Ex Intra-rater	5 studies, n = 296 [24,26,27,29,32]	0.77-0.99	100	Sufficient	High
Ex Inter-rater	2 studies, n = 113 [27,32]	0.87-0.97	100	Sufficient	Moderate

Abbreviations: Abd, abduction; Add, adduction; ER, external rotation; EX, extension; Flex, flexion; ICC, intraclass correlation coefficientIR, internal rotation.

TABLE 2 Summary of findings for measurement error of handheld dynamometry reported separately for each movement and for each type of reliability

Movement and type of reliability	Number of studies, number of participants	Summary of results: %MDC, range	Summary of results: proportion of %MDC ≤ 15%	Overall rating of results	Quality of evidence
IR Intra-rater	7 studies, n = 172[22-24,27,31,32,40]	0-29.0	40	Indeterminate	High
IR Inter-rater	5 studies, n = 186[22,23,27,32,40]	12.0-42.7	19	Insufficient	High
ER Intra-rater	7 studies, n = 185[22-24,31-33,40]	0-48.5	40	Indeterminate	High
ER Inter-rater	5 studies, n = 199[22,23,32,33,40]	15.0-33.3	6	Insufficient	High
Abd Intra-rater	3 studies, n = 68[28,32,40]	15.0-35.2	11	Insufficient	Moderate
Abd Inter-rater	3 studies, n = 144[28,32,40]	16.1-32.4	0	Insufficient	High
Add Intra-rater	1 study, n = 25[32]	51.0	0	Insufficient	Very low
Add Inter-rater	1 study, n = 101[32]	19.1-24.7	0	Insufficient	Low
Flex Intra-rater	4 studies, n = 83[24,28,32,40]	11.6-35.8	27	Indeterminate	Moderate
Flex Inter-rater	4 studies, n = 180[28,32,35,40]	16.1-39.9	0	Insufficient	High
Ex Intra-rater	4 studies, n = 65[24,26,27,32]	2.8-30.5	39	Indeterminate	Moderate
Ex Inter-rater	2 studies, n = 113[27,32]	8.6-25.2	50	Indeterminate	Moderate

Abbreviations: Abd, abduction; Add, adduction; ER, external rotation; EX, extension; Flex, flexion; IR, internal rotation; MDC_{95}, minimal detectable change.

TABLE 3 Summary of findings for construct validity of handheld dynamometry reported separately for handheld dynamometer compared with isokinetic dynamometer, externally fixed dynamometer, and spring scale dynamometer, and for subjects with symptoms compared to subjects without symptoms

	Number of studies, number of participants	Summary of results	Overall rating of results	Quality of evidence
HHD/ID	3 studies, n = 54[22,36,39]	*r*, range .28-.85	Insufficient	Very low
HHD/EFD	1 study, n = 20[37]	Mean diff., range (N) -6.5 to 29.9	Indeterminate	Low
HHD/spring scale	2 studies, n = 18[6,42]	*r*, range .77-.99	Sufficient	Very low
With/Without symptoms	1 study, n = 36[35]	*P*-value, range .89-.99	Insufficient	Very low

Abbreviations: diff., difference; EFD, externally fixed dynamometer; HHD, handheld dynamometer; ID, isokinetic dynamometer; N, Newton; *r*, Pearson's correlation coefficient.

"sufficient." The results of seven movements and type of reliability were rated as "insufficient" and the others as "indeterminate." The quality of evidence was graded as "high" or "moderate" except for adduction, which was "low" or "very low."

The sensitivity analysis showed that by setting the MIC to 10%, the results of all the movements and type of reliability were rated as "insufficient." By setting the MIC to 20%, the results were that two movements and type of reliability were rated as "sufficient," two as "insufficient," and eight as "indeterminate."

3.3 | Construct validity

Table 3 shows that the correlations between HHD and ID ranged from 0.28 to 0.85.[22,36,39] The results were rated as "insufficient," and the quality of evidence was "very low."

The comparison between HHD and externally fixed dynamometer showed "low"-quality evidence that two out of three mean differences were significantly different from zero.[37] The results were rated as "indeterminate."

The correlations between HHD and spring scale dynamometer were rated as "sufficient" (0.77-0.99),[6,42] but the quality of evidence was "very low."

One study compared subjects with and without shoulder symptoms[35] and found "very low" quality of evidence for no significant differences between groups.

4 | DISCUSSION

This systematic review included 28 studies investigating reliability, measurement error, or construct validity of shoulder muscle strength assessment with HHD. The methodological quality of the majority of studies was either adequate or doubtful assessed with the COSMIN risk of bias checklist.

Our results demonstrated that the reliability of HHD was overall sufficient with high or moderate quality of evidence for 11 out of 12 movements and type of reliability examined. The measurement error results were either insufficient or indeterminate with high or moderate quality of evidence for 10 out of 12 movements and type of reliability examined. Based on the evidence of low or very low quality, the validity of HHD when compared with other types of dynamometers was either sufficient, indeterminate, or insufficient, and HHD could not measure differences in muscle strength between groups with and without shoulder symptoms. No studies evaluating responsiveness were identified.

Reliability results indicate that HHD can be used to distinguish between individuals on group level. However, clinicians are often interested in the assessment of muscle strength in the same individual before and after an intervention. To be able to detect small but clinically relevant changes, the measurement error needs to be smaller than the MIC. For assessing MIC, anchor-based methods are preferred, as they include a definition of what is minimally important from a patient perspective. [50] Furthermore, MIC is dependent on the baseline score and thereby the study population and can be different for different groups of athletes, healthy non-athletes, and patients with shoulder disorders. [50] Torrens et al evaluated the MIC in the Constant score domains function and strength in patients treated with reverse shoulder arthroplasty and found the MIC for the strength domain to be 11.5 points. [51] Other studies have examined the MIC of the Constant score in patients with shoulder disorders, but did not specifically focus on the strength domain. [52,53] We found it appropriate to set the MIC at 15% and rate the measurement error results against this criterion. In addition, we conducted a sensitivity analysis by setting the MIC to 10% and 20%. Only a few of the included studies found measurement error results of MDC ≤15%. However, the results did not exceed the 75% limit and could not be rated as sufficient for any of the movements and type of reliability examined. Therefore, we must conclude that HHD cannot measure changes in muscle strength less than 15%, and it is also questionable if HHD can measure changes less than 20%. Whether this change is clinically meaningful depends on the expected magnitude of change in the population of interest. The expected change might differ substantially between athletes and patients. [50] For some athletes, a small change might be important even though it is unknown if it exceeds MDC. However, knowledge of the performance of the instrument will help athletes, patients, and testers to interpret the results of muscle strength assessment with HHD.

Our results of reliability and measurement error are similar to those reported in other reviews even though their inclusion criteria differed from ours. Schrama et al[1] found that 15 out of 38 studies showed acceptable intra-examiner reliability. However, their conclusions were based on a cutoff value of 0.90 for acceptable reliability and they only focused on reliability and not measurement error.

Our conclusion for convergent validity is in contrast with the conclusion reported by Stark et al[4] Based on similar results with correlation coefficients rating from 0.28 to 0.89, they concluded that HHD compared with ID can be regarded as a reliable and valid instrument to assess shoulder muscle strength.

4.1 | Strengths and limitations

To our knowledge, this is the first systematic review to summarize available research on measurement properties of HHD when used to assess shoulder muscle strength. It is considered a strength of the study to use the COSMIN methodology to evaluate the measurement properties of HHD. Furthermore, the detailed description of all steps increases the transparency of the methodological process. However, some limitations must be noted. Although we made a thorough literature search, relevant studies may have been missed. Additionally, using other quality assessment tools might have generated different conclusions.

5 | CONCLUSION

We conclude that the reliability (relative reliability) of HHD was overall sufficient and HHD can be used to distinguish individuals on the group level. The measurement error (absolute reliability) was not sufficient. HHD cannot measure changes less than 15%, and it is questionable if HHD can measure changes less than 20%. The conclusion is based primarily on high or moderate quality of evidence.

5.1 | Perspective

HHD is a frequently used method to objectively assess shoulder muscle strength. However, the results of this review indicate that while the method is useful to distinguish individuals on the group level, like when comparing the muscle strength in two or more groups of subjects, the evaluation of treatment effect on the individual level should be interpreted with caution. Changes in shoulder muscle strength in individuals should exceed the measurement error to be detected with HHD. Whether a 20% change seems realistic depends on the clinical context.

CONFLICT OF INTERESTS

The authors report no conflicts of interest.

ORCID

Lotte Sørensen https://orcid.org/0000-0001-5123-151X

REFERENCES

1. Schrama PPM, Stenneberg MS, Lucas C, Van Trijffel E. Intraexaminer reliability of hand-held dynamometry in the upper extremity: a systematic review. *Arch Phys Med Rehabil*. 2014;95(12):2444-2469.
2. Rabelo M, Nunes GS, da Costa Amante NM, de Noronha M, Fachin-Martins E. Reliability of muscle strength assessment in chronic post-stroke hemiparesis: a systematic review and meta-analysis. *Top Stroke Rehabil*. 2016;23(1):26-35.
3. Chamorro C, Armijo-Olivo S, De la Fuente C, Fuentes J, Javier Chirosa L. Absolute reliability and concurrent validity of hand held dynamometry and isokinetic dynamometry in the hip, knee and ankle joint: systematic review and meta-analysis. *Open Med*. 2017;12(1):359-375.
4. 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.
5. Baschung Pfister P, de Bruin ED, Sterkele I, Maurer B, de Bie RA, Knols RH. Manual muscle testing and hand-held dynamometry in people with inflammatory myopathy: An intra- and interrater reliability and validity study. *PLoS One*. 2018;13(3):e0194531.
6. Hayes K, Walton JR, Szomor ZL, Murrell GAC. Reliability of 3 methods for assessing shoulder strength. *J Shoulder Elbow Surg*. 2002;11(1):33-39.
7. Noreau L, Vachon J. Comparison of three methods to assess muscular strength in individuals with spinal cord injury. *Spinal Cord*. 1998;36(10):716-723.
8. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30(1):1-15.
9. Mokkin LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet HC, Terwee CB. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manuel; 2018; https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018-1.pdf. Accessed February 1, 2019
10. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19(4):539-549.
11. Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, De Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21(4):651-657.
12. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the sem. *J Strength Cond Res*. 2005;19(1):231-240.
13. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59(10):1033-1039.
14. Kottner J, Gajewski BJ, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). *Int J Nurs Stud*. 2011;48(6):659-660.
15. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.
16. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1.
17. Higgins JPTGSE. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. http://www.handbook.cochrane.org. Accessed March 3, 2019
18. Schünemann H. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach; 2013. https://gdt.gradepro.org/app/handbook/handbook.html. Accessed March 3, 2019
19. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123.
20. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171-1179.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42.
22. Holt KL, Raper DP, Boettcher CE, Waddington GS, Drew MK. Hand-held dynamometry strength measures for internal and external rotation demonstrate superior reliability, lower minimal detectable change and higher correlation to isokinetic dynamometry than externally-fixed dynamometry of the shoulder. *Phys Ther Sport*. 2016;21:75-81.
23. 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.
24. McLaine SJ, Ginn KA, Kitic CM, Fell JW, Bird ML. The reliability of strength tests performed in elevated shoulder positions using a handheld dynamometer. *J Sport Rehabil*. 2016.25(2):1–4.
25. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.

26. Awatani T, Mori S, Shinohara J, et al. Same-session and between-day intra-rater reliability of hand-held dynamometer measurements of isometric shoulder extensor strength. *J Phys Ther Sci*. 2016;28(3):936-939.
27. Awatani T, Morikita I, Shinohara J, et al. Intra- and inter-rater reliability of isometric shoulder extensor and internal rotator strength measurements performed using a hand-held dynamometer. *J Phys Ther Sci*. 2016;28(11):3054-3059.
28. Balogun JA, Powell R, Trullender B, Olson S, Balogun AO. Intra-and inter-tester reliability of the Nicholas?? hand-held dynamometer during evaluation of upper extremity isometric muscle strength. *Eur J Phys Med Rehabil*. 1998;8(2):48-53.
29. Bohannon RW. Reference values for extremity muscle strength obtained by hand-held dynamometry from adults aged 20 to 79 years. *Arch Phys Med Rehabil*. 1997;78(1):26-32.
30. Byl NN, Richards S, Asturias J. Intrarater and interrater reliability of strength measurements of the biceps and deltoid using a hand held dynamometer. *J Orthop Sports Phys Ther*. 1988;9(12):395-398.
31. 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.
32. Fieseler G, Laudner KG, Irlenbusch L, et al. Inter- and intrarater reliability of goniometry and hand held dynamometry for patients with subacromial impingement syndrome. *J Exerc Rehabil*. 2017;13(6):704-710.
33. 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.
34. Leggin BG, Neuman RM, Iannotti JP, Williams GR, Thompson EC. Intrarater and interrater reliability of three isometric dynamometers in assessing shoulder strength. *J Shoulder Elbow Surg*. 1996;5(1):18-24.
35. Sciascia A, Uhl T. Reliability of strength and performance testing measures and their ability to differentiate persons with and without shoulder symptoms. *Int J Sports Phys Ther*. 2015;10(5):655-666.
36. Sullivan SJ, Chesley A, Hebert G, McFaull S, Scullion D. The validity and reliability of hand-held dynamometry in assessing isometric external rotator performance. *J Orthop Sports Phys Ther*. 1988;10(6):213-217.
37. Vermeulen HM, de Bock GH, van Houwelingen HC, et al. A comparison of two portable dynamometers in the assessment of shoulder and elbow strength. *Physiotherapy*. 2005;91(2):101-112.
38. Beshay N, Lam PH, Murrell GAC. Assessing the reliability of shoulder strength measurement: hand-held versus fixed dynamometry. *Shoulder Elb*. 2011;3(4):244-251.
39. Burnham SR, Bell CG, Olenik CL, Reid CD. Shoulder abduction strength measurement in football players: reliability and validity of two field tests. *Clin J Sport Med*. 1995;5(2):90-94.
40. Dollings H, Sandford F, O'Conaire E, Lewis JS. Shoulder strength testing: the intra- and inter-tester reliability of routine clinical tests, using the PowerTrack™ II Commander. *Shoulder Elb*. 2012;4(2):131-140.
41. Donatelli R, Ellenbecker TS, Ekedahl SR, Wilkes JS, Kocher K, Adam J. Assessment of shoulder strength in professional baseball pitchers. *J Orthop Sports Phys Ther*. 2000;30(9):544-551.
42. Johansson KM, Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment. *J Shoulder Elbow Surg*. 2005;14(3):273-278.
43. Magnusson SP, Gleim GW, Nicholas JA. Subject variability of shoulder abduction strength testing. *Am J Sports Med*. 1990;18(4):349-353.
44. McMahon LM, Burdett R, Whitney SL. Effects of muscle group and placement site on reliability of hand-held dynamometry strength measurements. *J Orthop Sports Phys Ther*. 1992;15(5):236-242.
45. Ottenbacher KJ, Branch LG, Ray L, Gonzales VA, Peek MK, Hinman MR. The reliability of upper- and lower-extremity strength testing in a community survey of older adults. *Arch Phys Med Rehabil*. 2002;83(10):1423-1427.
46. Phillips BA, Lo SK, Mastaglia FL. Muscle force measured using "break" testing with a hand-held myometer in normal subjects aged 20 to 69 years. *Arch Phys Med Rehabil*. 2000;81(5):653-661.
47. Riemann BL, Davies GJ, Ludwig L, Gardenhour H. Hand-held dynamometer testing of the internal and external rotator musculature based on selected positions to establish normative data and unilateral ratios. *J Shoulder Elbow Surg*. 2010;19(8):1175-1183.
48. Cadogan A, Laslett M, Hing W, McNair P, Williams M. Reliability of a new hand-held dynamometer in measuring shoulder range of motion and strength. *Man Ther*. 2011;16(1):97-101.
49. O'Shea SD, Taylor NF, Paratz JD. Measuring muscle strength for people with chronic obstructive pulmonary disease: retest reliability of hand-held dynamometry. *Arch Phys Med Rehabil*. 2007;88(1):32-36.
50. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes*. 2006;4:54.
51. Torrens C, Guirro P, Santana F. The minimal clinically important difference for function and strength in patients undergoing reverse shoulder arthroplasty. *J Shoulder Elbow Surg*. 2016;25(2):262-268.
52. Holmgren T, Oberg B, Adolfsson L, Björnsson Hallgren H, Johansson K. Minimal important changes in the Constant-Murley score in patients with subacromial pain. *J Shoulder Elbow Surg*. 2014;23(8):1083-1090.
53. Kukkonen J, Kauko T, Vahlberg T, Joukainen A, Äärimaa V. Investigating minimal clinically important difference for Constant score in patients undergoing rotator cuff surgery. *J Shoulder Elbow Surg*. 2013;22(12):1650-1655.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Sørensen L, Oestergaard LG, van Tulder M, Petersen AK. Measurement properties of handheld dynamometry for assessment of shoulder muscle strength: A systematic review. *Scand J Med Sci Sports*. 2020;30:2305–2328. https://doi.org/10.1111/sms.13805

Appendix 1

Search strategy in PubMed

#1	Shoulder [MeSH] OR Shoulder Pain [MeSH] OR Shoulder Impingement Syndrome[MeSH] OR Shoulder Joint [MeSH] OR shoulder[tiab] OR shoulders[tiab] OR rotator[tiab] OR rotators[tiab]
#2	Muscle Strength Dynamometer [MeSH] OR Muscle Strength [MeSH] OR Muscle Contraction [MeSH] OR strength[tiab] OR force[tiab] OR power[tiab] OR dynamometer[tiab] OR dynamometry[tiab] OR hand-held[tiab] OR handheld[tiab] OR isokinetic[tiab] OR isometric[tiab] OR peak torque[tiab] OR EMG[tiab]
#3	(instrumentation[sh] OR methods[sh] OR Validation Studies[pt] OR Comparative Study[pt] OR ‘‘psychometrics’’ [MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR ‘‘outcome assessment (health care)’’[MeSH] OR outcome assessment[tiab] OR outcome measure*[tw] OR ‘‘observer variation’’[MeSH] OR observer variation[tiab] OR ‘‘Health Status Indicators’’[Mesh] OR ‘‘reproducibility of results’’[MeSH] OR reproducib*[tiab] OR ‘‘discriminant analysis’’[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR ‘‘internal consistency’’[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tiab] OR precision[tiab] OR imprecision[tiab] OR ‘‘precise values’’[tiab] OR test– retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab* [tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant [tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa’s[tiab] OR kappas[tiab] OR repeatab*[tiab] OR ((replicab*[tiab] OR repeated[tiab]) AND (measure[tiab] OR measures[tiab] OR findings[tiab] OR result[tiab] OR results[tiab] OR test[tiab] OR tests[tiab])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR ‘‘known group’’[tiab] OR factor analysis[tiab] OR factor analyses[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR item discriminant[tiab] OR interscale correlation*[tiab] OR error[tiab] OR errors[tiab] OR ‘‘individual variability’’[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR ‘‘standard error of measurement’’[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR meaningful change [tiab] OR ‘‘ceiling effect’’[tiab] OR ‘‘floor effect’’[tiab] OR ‘‘Item response model’’[tiab] OR IRT[tiab] OR Rasch[tiab] OR ‘‘Differential item functioning’’[tiab] OR DIF[tiab] OR ‘‘computer adaptive testing’’[tiab] OR ‘‘item bank’’[tiab] OR ‘‘cross-cultural equivalence’’[tiab])
#4	#1 AND #2 AND #3

Appendix 2

Results of reliability and measurement error for handheld dynamometry

Author/year	Movement	Type of reliability	Test condition	Reliability ICC_{model} (95% CI)/*r*	Measurement error SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
Awatani[a], 2016	Ex	Intra	WS, day 1, dom. WS, day 1, non-dom. WS, day 2, dom. WS, day 2, non-dom. WS, day 1, dom. WS, day 1, non-dom. WS, day 2, dom. WS, day 2, non-dom. BS, dom. BS, non-dom. BS, dom. BS, non-dom.	$ICC_{1,1}$ = 0.93 (0.83-0.98) $ICC_{1,1}$ = 0.85 (0.68-0.95) $ICC_{1,1}$ = 0.88 (0.73-0.96) $ICC_{1,1}$ = 0.94 (0.87-0.98) $ICC_{1,3}$ = 0.98 (0.94-0.99) $ICC_{1,3}$ = 0.95 (0.86-0.98) $ICC_{1,3}$ = 0.96 (0.89-0.99) $ICC_{1,3}$ = 0.98 (0.95-0.99) $ICC_{1,1}$ = 0.90 (0.68-0.97) $ICC_{1,1}$ = 0.89 (0.66-0.97) $ICC_{1,2}$ = 0.95 (0.81-0.99) $ICC_{1,2}$ = 0.94 (0.80-0.98)	8.1 9.2 8.7 5.5 4.8 5.6 5.2 4.0 8.6 7.6 6.2 5.5		22.4 (1) 25.6 (1) 24.0 (1) 15.2 (1) 13.3 (1) 15.6 (1) 14.5 (1) 10.9 (1) 23.8 (1) 20.9 (1) 17.3 (1) 15.2 (1)	23.2 (c) 30.7 (c) 24.5 (c) 17.0 (c) 13.8 (c) 18.7 (c) 14.8 (c) 12.2 (c) 24.5 (c) 24.2 (c) 17.8 (c) 17.6 (c)
Awatani[b], 2016	IR	Intra	Day 1, examiner 1, dom. Day 1, examiner 1, non-dom. Day 1, examiner 2, dom. Day 1, examiner 2, non-dom. Day 2, examiner 1, dom. Day 2, examiner 1, non-dom. Day 2, examiner 2, dom. Day 2, examiner 2, non-dom. Examiner 1, dom. Examiner 1, non-dom. Examiner 2, dom. Examiner 2, non-dom.	$ICC_{1,3}$ = 0.98 (0.94-0.99) $ICC_{1,3}$ = 0.93 (0.81-0.98) $ICC_{1,3}$ = 0.97 (0.92-0.99) $ICC_{1,3}$ = 0.97 (0.83-0.99) $ICC_{1,3}$ = 0.94 (0.85-0.98) $ICC_{1,3}$ = 0.91 (0.76-0.97) $ICC_{1,3}$ = 0.92 (0.80-0.98) $ICC_{1,3}$ = 0.90 (0.75-0.97) $ICC_{1,2}$ = 0.94 (0.81-0.98) $ICC_{1,2}$ = 0.95 (0.84-0.99) $ICC_{1,2}$ = 0.96 (0.87-0.99) $ICC_{1,2}$ = 0.96 (0.86-0.99)	4 2 5.2 4.6 4.3 5.1 5.6 6.6 4.3 5.6 3.8 4.7 3.9	5.4 7.8 5.5 6.2 6.4 8.2 8.6 6.5 7.1 5.6 5.8 5.7	11.8 (1) 14.5 (1) 12.6 (1) 12.0 (1) 14.2 (1) 15.6 (1) 18.4 (1) 11.9 (1) 15.6 (1) 10.5 (1) 13.0 (1) 10.7 (1)	15.0 (1) 21.7 (1) 5.2 (1) 17.1 (1) 17.9 (1) 22.8 (1) 23.9 (1) 17.9 (1) 19.8 (1) 15.6 (1) 16.2 (1) 15.8 (1)

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability ICC_{model} (95% CI)/*r*	Measurement error SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
		Inter	Day 1, dom.	$ICC_{2,2}$ = 0.97 (0.88-0.99)	4.5	5.6	12.4 (1)	15.4 (1)
			Day 1, non-dom.	$ICC_{2,2}$ = 0.95 (0.84-0.99)	4.1	6.0	11.4 (1)	16.6 (1)
			Day 2, dom.	$ICC_{2,2}$ = 0.97 (0.91-0.99)	3.4	4.3	9.4 (1)	12.0 (1)
			Day 2, non-dom.	$ICC_{2,2}$ = 0.96 (0.85-0.99)	3.7	5.4	10.7 (1)	15.1 (1)
	Ex	Intra	Day 1, examiner 1, dom.	$ICC_{1,3}$ = 0.95 (0.88-0.99)	4.1	4.8	11.4 (1)	11.3 (1)
			Day 1, examiner 1, non-dom.	$ICC_{1,3}$ = 0.88 (0.69-0.96)	4.8	6.4	13.3 (1)	17.6 (1)
			Day 1, examiner 2, dom.	$ICC_{1,3}$ = 0.96 (0.89-0.99)	3.4	4.0	9.3 (1)	11.0 (1)
			Day 1, examiner 2, non-dom.	$ICC_{1,3}$ = 0.87 (0.66-0.96)	5.1	6.6	14.1 (1)	18.2 (1)
			Day 2, examiner 1, dom.	$ICC_{1,3}$ = 0.83 (0.57-0.95)	7.1	8.7	19.5 (1)	24.1 (1)
			Day 2, examiner 1, non-dom.	$ICC_{1,3}$ = 0.87 (0.66-0.96)	5.6	7.4	15.6 (1)	20.6 (1)
			Day 2, examiner 2, dom.	$ICC_{1,3}$ = 0.81 (0.52-0.94)	7.1	8.8	19.5 (1)	24.3 (1)
			Day 2, examiner 2, non-dom.	$ICC_{1,3}$ = 0.86 (0.63-0.96)	5.9	8.2	16.3 (1)	22.7 (1)
			Examiner 1, dom.	$ICC_{1,2}$ = 0.88 (0.62-0.97)	5.7	6.8	15.7 (1)	18.8 (1)
			Examiner 1, non-dom.	$ICC_{1,2}$ = 0.92 (0.73-0.98)	3.8	5.0	10.5 (1)	13.9 (1)
			Examiner 2, dom.	$ICC_{1,2}$ = 0.77 (0.24-0.93)	7.1	8.6	19.7 (1)	23.9 (1)
			Examiner 2, non-dom.	$ICC_{1,2}$ = 0.85 (0.49-0.96)	5.2	7.0	14.5 (1)	19.5 v
		Inter	Day 1, dom.	$ICC_{2,2}$ = 0.93 (0.75-0.98)	4.6	5.4	12.7 (1)	14.9 (1)
			Day 1, non-dom.	$ICC_{2,2}$ = 0.96 (0.88-0.99)	2.4	3.1	6.6 (1)	8.6 (1)
			Day 2, dom.	$ICC_{2,2}$ = 0.93 (0.74-0.98)	3.9	4.9	10.9 (1)	13.4 (1)
			Day 2, non-dom.	$ICC_{2,2}$ = 0.87 (0.57-0.96)	5.0	6.8	14.0 (1)	18.9 (1)
Balogun, 1998	Abd	Intra	WS, examiner 1	$ICC_{1,1}$ = 0.95	6.96 (c)		19.3 (c)	22.2 (c)
			WS, examiner 2	$ICC_{1,1}$ = 0.95	6.47 (c)		17.9 (c)	21.1 (c)
			BS, examiner 1	$ICC_{1,1}$ = 0.91	8.63 (c)		23.9 (c)	27.4 (c)
			BS, examiner 2	$ICC_{1,1}$ = 0.96	5.69 (c)		15.8 (c)	18.3 (c)
		Inter	WS	$ICC_{1,1}$ = 0.95	6.67 (c)		18.5 (c)	21.1 (c)
			BS	$ICC_{1,1}$ = 0.97	5.10 (c)		14.1 (c)	16.1 (c)
	Flex	Intra	WS, examiner 1	$ICC_{1,1}$ = 0.95	6.28 (c)		17.4 (c)	20.5 (c)
			WS, examiner 2	$ICC_{1,1}$ = 0.97	5.20 (c)		14.4 (c)	16.9 (c)
			BS, examiner 1	$ICC_{1,1}$ = 0.95	5.98 (c)		16.6 (c)	19.7 (c)
			BS, examiner 2	$ICC_{1,1}$ = 0.95	6.08 (c)		16.9 (c)	19.7 (c)

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability: ICC_{model} (95% CI)/*r*	Measurement error: SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
		Inter	WS BS	$ICC_{1,1}$ = 0.96 $ICC_{1,1}$ = 0.97	5.78 (c) 5.00 (c)		16.0 (c) 13.9 (c)	19.1 (c) 16.1 (c)
Beshay, 2011	IR	Intra	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.98 $ICC_{2,1}$ = 0.94				
		Inter	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.89 $ICC_{2,1}$ = 0.97				
	ER	Intra	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.95 $ICC_{2,1}$ = 0.95				
		Inter	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.89 $ICC_{2,1}$ = 0.98				
	Abd	Intra	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.93 $ICC_{2,1}$ = 0.96				
		Inter	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.83 $ICC_{2,1}$ = 0.97				
	Add	Intra	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.98 $ICC_{2,1}$ = 0.97				
		Inter	Asymptomatic Symptomatic	$ICC_{2,1}$ = 0.92 $ICC_{2,1}$ = 0.86				
Bohannon, 1997	ER	Intra	Dominant Non-dominant	$ICC_{3,1}$ = 0.97 $ICC_{3,1}$ = 0.95			8.0 (3) 9.8 (3)	
	Abd	Intra	Dominant Non-dominant	$ICC_{3,1}$ = 0.97 $ICC_{3,1}$ = 0.96			11.7 (3) 13.1 (3)	
	Ex	Intra	Dominant Non-dominant	$ICC_{3,1}$ = 0.97 $ICC_{3,1}$ = 0.97			16.3 (3) 15.9 (3)	
Burnham, 1995	Abd	Intra	WS, rested WS, competition BS, rested BS, competition	$ICC_{?}$ = 0.60 $ICC_{?}$ = 0.81 $ICC_{?}$ = 0.55 $ICC_{?}$ = 0.70				
Byl, 1988	Abd	Intra	Without stab., trial 1/2 Without stab., trial 1/3 Without stab., trial 2/3 With stab., trial 1/2 With stab., trial 1/3 With stab., trial 2/3	*r* = .83 *r* = .84 *r* = .89 *r* = .92 *r* = .94 *r* = .94				
		Inter	Without stab. With stab.	*r* = .67 *r* = .84				

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability ICC_{model} (95% CI)/*r*	Measurement error SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = %MDC_{95} (c) = %MDC_{95} calculated
Cadogan, 2010	ER	Intra	Examiner 1, affected arm	$ICC_{2,1}$ = 0.96 (0.94-0.98)			1.7 (4) 1.9 (4)	
			Examiner 1, unaffected arm	$ICC_{2,1}$ = 0.94 (0.90-0.97)			3.2 (4) 3.0 (4)	
			Examiner 2, affected arm	$ICC_{2,1}$ = 0.91 (0.85-0.95)				
			Examiner 2, unaffected arm	$ICC_{2,1}$ = 0.93 (0.87-0.96)				
		Inter	First trial, affected arm	$ICC_{2,1}$ = 0.69 (0.23-0.87)			4.4 (4) 3.6 (4)	
			First trial, unaffected arm	$ICC_{2,1}$ = 0.68 (−0.05-0.89)			4.0 (4) 3.2 (4)	
			Mean of 3, affected arm	$ICC_{2,1}$ = 0.74 (0.36-0.89)				
			Mean of 3, unaffected arm	$ICC_{2,1}$ = 0.70 (−0.04-0.90)				
	Abd	Intra	Examiner 1, affected arm	$ICC_{2,1}$ = 0.91 (0.85-0.95)			7.0 (4) 4.5 (4)	
			Examiner 1, unaffected arm	$ICC_{2,1}$ = 0.95 (0.91-0.97)			3.6 (4) 3.8 (4)	
			Examiner 2, affected arm	$ICC_{2,1}$ = 0.95 (0.92-0.98)				
			Examiner 2, unaffected arm	$ICC_{2,1}$ = 0.95 (0.92-0.97)				
		Inter	First trial, affected arm	$ICC_{2,1}$ = 0.81 (0.42-0.92)			7.3 (4) 8.5 (4)	
			First trial, unaffected arm	$ICC_{2,1}$ = 0.77 (0.49-0.89)			6.3 (4) 7.6 (4)	
			Mean of 3, affected arm	$ICC_{2,1}$ = 0.84 (0.54-0.93)				
			Mean of 3, unaffected arm	$ICC_{2,1}$ = 0.77 (0.46-0.89)				
Cools, 2014	IR	Intra	Seated, 0° abd, examiner 1	$ICC_{3,1}$ = 0.99 (0.97-0.99)	6.03 9.48		14.07 (2) 22.11 (2)	11.1 (c) 17.2 (c)
			Seated, 0° abd, examiner 2	$ICC_{3,1}$ = 0.96 (0.93-0.98)	5.37 7.22		12.53 (2) 16.84 (2)	12.8 (c) 17.2 (c)
			Seated, 90° abd, examiner 1	$ICC_{3,1}$ = 0.99 (0.98-0.99)	4.11 5.40		9.58 (2) 12.60 (2)	8.9 (c) 11.6 (c)
			Seated, 90° abd, examiner 2	$ICC_{3,1}$ = 0.98 (0.96-0.99)	4.09 6.67		9.55 (2) 15.56 (2)	10.3 (c) 16.1 (c)
			Supine, 0° abd, examiner 1	$ICC_{3,1}$ = 0.99 (0.98-0.99)	4.88 6.09		11.39 (2) 14.21 (2)	10.5 (c) 13.9 (c)
			Supine, 0° abd examiner 2	$ICC_{3,1}$ = 0.98 (0.96-0.99)				
			Supine, 90° abd, examiner 1	$ICC_{3,1}$ = 0.99 (0.98-0.99)				
			Supine, 90° abd, examiner 2	$ICC_{3,1}$ = 0.98 (0.97-0.99)				
			Prone, 90° abd, examiner 1	$ICC_{3,1}$ = 0.99 (0.98-0.99)				
			Prone, 90° abd, examiner 2	$ICC_{3,1}$ = 0.98 (0.97-0.99)				

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability: ICC_{model} (95% CI)/*r*	Measurement error: SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = %MDC_{95} (c) = %MDC_{95} calculated
		Inter	Seated, 0° abd	$ICC_{2,k}$ = 0.98	7.76		21.50 (2)	14.1 (c)
			Seated, 90° abd	(0.95-0.99)	7.57		20.99 (2)	18.0 (c)
			Supine, 0° abd	$ICC_{2,k}$ = 0.98	7.56		20.96 (2)	16.4 (c)
			Supine, 90° abd	(0.96-0.99)	5.98		16.57 (2)	14.7 (c)
			Prone, 90° abd	$ICC_{2,k}$ = 0.96	9.60		26.60 (2)	21.1 (c)
				(0.91-0.98)				
				$ICC_{2,k}$ = 0.99				
				(0.97-0.99)				
				$ICC_{2,k}$ = 0.97				
				(0.92-0.98)				
	ER	Intra	Seated, 0° abd,	$ICC_{3,1}$ = 0.97	4.94		11.52 (2)	13.9 (c)
			examiner 1	(0.95-0.99)	5.48		12.80 (2)	16.1 (c)
			Seated, 0° abd,	$ICC_{3,1}$ = 0.96	4.59		10.72 (2)	16.6 (c)
			examiner 2	(0.93-0.98)	6.06		14.14 (2)	23.0 (c)
			Seated, 90° abd,	$ICC_{3,1}$ = 0.95	4.30		10.04 (2)	10.5 (c)
			examiner 1	(0.91-0.98)	7.13		16.63 (2)	18.3 (c)
			Seated, 90° abd,	$ICC_{3,1}$ = 0.93	6.01		14.03 (2)	20.5 (c)
			examiner 2	(0.88-0.97)	5.41		12.62 (2)	19.4 (c)
			Supine, 0° abd,	$ICC_{3,1}$ = 0.98	3.37		7.87 (2)	14.1 (c)
			examiner 1	(0.97-0.99)	4.75		11.09 (2)	20.2 (c)
			Supine, 0° abd	$ICC_{3,1}$ = 0.95				
			examiner 2	(0.91-0.97)				
			Supine, 90° abd,	$ICC_{3,1}$ = 0.95				
			examiner 1	(0.91-0.97)				
			Supine, 90° abd,	$ICC_{3,1}$ = 0.95				
			examiner 2	(0.91-0.98)				
			Prone, 90° abd,	$ICC_{3,1}$ = 0.97				
			examiner 1	(0.94-0.98)				
			Prone, 90° abd,	$ICC_{3,1}$ = 0.94				
			examiner 2	(0.89-0.96)				
		Inter	Seated, 0° abd	$ICC_{2,k}$ = 0.96	6.00		16.63 (2)	17.2 (c)
			Seated, 90° abd	(0.92-0.98)	4.94		13.70 (2)	18.3 (c)
			Supine, 0° abd	$ICC_{2,k}$ = 0.94	7.13		19.77 (2)	18.0 (c)
			Supine, 90° abd	(0.88-0.97)	6.36		17.62 (2)	22.2 (c)
			Prone, 90° abd	$ICC_{2,k}$ = 0.95	3.54		9.82 (2)	15.0 (c)
				(0.89-0.98)				
				$ICC_{2,k}$ = 0.94				
				(0.87-0.97)				
				$ICC_{2,k}$ = 0.96				
				(0.93-0.98)				
Dollings, 2012	IR	Intra	Examiner 1, right	$ICC_{2,3}$ = 0.97	8.5		23.6 (c)	20.5 (c)
			Examiner 1, left	(0.92-0.99)	9.3		25.8 (c)	20.8 (c)
			Examiner 2, right	$ICC_{2,3}$ = 0.98	6.7		18.6 (c)	14.7 (c)
			Examiner 2, left	(0.96-0.99)	8.5		23.6 (c)	20.0 (c)
				$ICC_{2,3}$ = 0.98				
				(0.97-0.99)				
				$ICC_{2,3}$ = 0.97				
				(0.94-0.99)				
		Inter	Right	$ICC_{2,3}$ = 0.93	12.9		35.8 (c)	29.7 (c)
			Left	(0.57-0.98)	14.7		40.7 (c)	33.5 (c)
				$ICC_{2,3}$ = 0.94				
				(0.86-0.97)				

(Continues)

Appendix 2 (Continued)

				Reliability	Measurement error			
Author/year	Movement	Type of reliability	Test condition	ICC_{model} (95% CI)/*r*	SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = %MDC_{95} (c) = %MDC_{95} calculated
	ER	Intra	Examiner 1, right Examiner 1, left Examiner 2, right Examiner 2, left	$ICC_{2,3}$ = 0.97 (0.93-0.99) $ICC_{2,3}$ = 0.98 (0.96-0.99) $ICC_{2,3}$ = 0.96 (0.90-0.98) $ICC_{2,3}$ = 0.95 (0.87-0.98)	6.2 6.7 7.1 7.6		17.2 (c) 18.6 (c) 19.7 (c) 21.1 (c)	18.8 (c) 20.2 (c) 20.5 (c) 20.8 (c)
		Inter	Right Left	$ICC_{2,3}$ = 0.92 (0.81-0.97) $ICC_{2,3}$ = 0.94 (0.82-9.98)	10.2 9.8		28.3 (c) 27.2 (c)	29.9 (c) 28.0 (c)
	Abd	Intra	Examiner 1, right Examiner 1, left Examiner 2, right Examiner 2, left	$ICC_{2,3}$ = 0.94 (0.86-0.98) $ICC_{2,3}$ = 0.94 (0.86-0.98) $ICC_{2,3}$ = 0.93 (0.85-0.97) $ICC_{2,3}$ = 0.92 (0.81-0.97)	8.5 9.3 7.6 9.8		23.6 (c) 25.8 (c) 21.1 (c) 27.2 (c)	19.7 (c) 23.6 (c) 15.0 (c) 21.6 (c)
		Inter	Right Left	$ICC_{2,3}$ = 0.77 (0.45-0.91) $ICC_{2,3}$ = 0.87 (0.63-0.95)	15.1 13.8		41.9 (c) 38.3 (c)	32.4 (c) 32.4 (c)
	Flex	Intra	Examiner 1, right Examiner 1, left Examiner 2, right Examiner 2, left	$ICC_{2,3}$ = 0.97 (0.89-0.98) $ICC_{2,3}$ = 0.89 (0.62-0.92) $ICC_{2,3}$ = 0.89 (0.74-0.95) $ICC_{2,3}$ = 0.92 (0.82-0.97)	8.5 15.6 12.5 11.6		23.6 (c) 43.2 (c) 34.6 (c) 32.2 (c)	18.3 (c) 35.8 (c) 27.2 (c) 26.1 (c)
		Inter	Right Left	$ICC_{2,3}$ = 0.91 (0.79-0.96) $ICC_{2,3}$ = 0.92 (0.82-0.97)	12.9 11.6		35.8 (c) 32.2 (c)	28.0 (c) 26.3 (c)
Donatelli, 2000	IR	Intra	Plane of scapula 90° abd	$ICC_{2,k}$ = 0.82 $ICC_{2,k}$ = 0.93				
	ER	Intra	Plane of scapula 90° abd	$ICC_{2,k}$ = 0.82 $ICC_{2,k}$ = 0.96				
Fieseler, 2015	IR	Intra	Healthy, throwing arm Healthy, non-throwing arm Handball player, throwing arm Handball player, non-throwing arm	$ICC_{?}$ = 0.99 (0.99-1.00) $ICC_{?}$ = 1.00 (0.99-1.00) $ICC_{?}$ = 0.94 (0.87-0.98) $ICC_{?}$ = 0.95 (0.89-0.98)	3.59 0 4.76 3.63		-10.9-11.4 (5) -6.6-9.7 (5) -17.0-19.4 (5) -14.2-14.0 (5)	10.3 (c) (c) 14.4 (c) 11.9 (c)

(Continues)

Appendix 2 (Continued)

				Reliability	Measurement error			
Author/year	Movement	Type of reliability	Test condition	ICC_{model} (95% CI)/*r*	SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
	ER	Intra	Healthy, throwing arm Healthy, non-throwing arm Handball player, throwing arm Handball player, non-throwing arm	$ICC_{?}$ = 0.99 (0.97-1.00) $ICC_{?}$ = 1.00 (0.99-1.00) $ICC_{?}$ = 0.95 (0.89-0.98) $ICC_{?}$ = 0.96 (0.91-0.98)	4.48 0 4.27 4.29		-6.8-15.4 (5) -11.8-11.6 (5) -18.0-15.0 (5) -18.5-14.1 (5)	10.5 (c) (c) 11.1 (c) 11.1 (c)
Fieseler, 2017	IR	Intra	Patients with SIS	$ICC_{?}$ = 0.97 (0.92-0.99)	4.83		13.4 (c)	21.9 (c)
		Inter	Patients with SIS Healthy subjects	$ICC_{?}$ = 0.93 (0.85-0.97) $ICC_{?}$ = 0.72 (0.55-0.82)	7.75 11.8		21.5 (c) 32.7 (c)	34.4 (c) 42.7 (c)
	ER	Intra	Patients with SIS	$ICC_{?}$ = 0.90 (0.78-0.96)	9.08		25.2 (c)	48.5 (c)
		Inter	Patients with SIS Healthy subjects	$ICC_{?}$ = 0.96 (0.88-0.98) $ICC_{?}$ = 0.96 (0.93-0.97)	5.75 9.41		15.9 (c) 26.1 (c)	29.4 (c) 27.4 (c)
	Abd	Intra	Patients with SIS	$ICC_{?}$ = 0.94 (0.86-0.97)	5.00		13.9 (c)	35.2 (c)
		Inter	Patients with SIS Healthy subjects	$ICC_{?}$ = 0.95 (0.89-0.98) $ICC_{?}$ = 0.91 (0.85-0.94)	4.65 5.69		12.9 (c) 15.8 (c)	32.4 (c) 24.7 (c)
	Add	Intra	Patients with SIS	$ICC_{?}$ = 0.87 (0.70-0.94)	9.86		27.3 (c)	51.0 (c)
		Inter	Patients with SIS Healthy subjects	$ICC_{?}$ = 0.98 (0.95-0.99) $ICC_{?}$ = 0.93 (0.89-0.96)	3.51 6.19		9.7 (c) 17.2 (c)	19.1 (c) 24.7 (c)
	Flex	Intra	Patients with SIS	$ICC_{?}$ = 0.94 (0.86-0.97)	4.36		12.1 (c)	26.1 (c)
		Inter	Patients with SIS Healthy subjects	$ICC_{?}$ = 0.89 (0.75-0.95) $ICC_{?}$ = 0.95 (0.92-0.97)	7.03 4.18		19.5 (c) 11.6 (c)	39.9 (c) 16.6 (c)
	Ex	Intra	Patients with SIS	$ICC_{?}$ = 0.94 (0.86-0.98)	7.96		22.1 (c)	33.0 (c)
		Inter	Patients with SIS Healthy subjects	$ICC_{?}$ = 0.97 (0.94-0.99) $ICC_{?}$ = 0.96 (0.93-0.97)	6.33 5.48		17.5 (c) 15.2 (c)	25.2 (c) 18.3 (c)
Hayes, 2002	IR	Intra		$ICC_{2,1}$ = 0.85 (0.64-0.96)				
		Inter		$ICC_{2,1}$ = 0.85 (0.62-0.97)				

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability ICC_{model} (95% CI)/*r*	Measurement error SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = %MDC_{95} (c) = %MDC_{95} calculated
	ER	Intra		$ICC_{2,1}$ = 0.92 (0.78-0.98)				
		Inter		$ICC_{2,1}$ = 0.82 (0.55-0.96)				
	Flex	Intra		$ICC_{2,1}$ = 0.96 (0.88-0.99)				
		Inter		$ICC_{2,1}$ = 0.92 (0.75-0.99)				
Holt, 2016	IR	Intra	Highest value, EFD, left Highest value, EFD, right Highest value, HHD, left Highest value, HHD, right Average value, EFD, left Average value, EFD, right Average value, HHD, left Average value, HHD, right	$ICC_{3,1}$ = 0.96 (0.91-0.99) $ICC_{3,1}$ = 0.91 (0.88-0.98) $ICC_{3,1}$ = 0.96 (0.90-0.98) $ICC_{3,1}$ = 0.95 (0.89-0.98) $ICC_{3,1}$ = 0.96 (0.90-0.98) $ICC_{3,1}$ = 0.89 (0.73-0.95) $ICC_{3,1}$ = 0.96 (0.89-0.98) $ICC_{3,1}$ = 0.96 (0.90-0.98)	2.65 3.87 3.05 3.20 2.45 3.97 2.95 2.93	7.1 9.6 6.5 6.4 7.0 10.5 6.3 6.3	7.34 (1) 10.71 (1) 8.46 (1) 8.88 (1) 6.79 (1) 11.00 (1) 8.18 (1) 8.13 (1)	19.8 (1) 26.6 (1) 18.0 (1) 17.9 (1) 19.3 (1) 29.0 (1) 18.1 (1) 17.3 (1)
		Inter	Highest value, EFD, left Highest value, EFD, right Highest value, HHD, left Highest value, HHD, right Average value, EFD, left Average value, EFD, right Average value, HHD, left Average value, HHD, right	$ICC_{2,1}$ = 0.96 (0.91-0.99) $ICC_{2,1}$ = 0.91 (0.78-0.96) $ICC_{2,1}$ = 0.96 (0.90-0.98) $ICC_{2,1}$ = 0.95 (0.89-0.98) $ICC_{2,1}$ = 0.96 (0.91-0.98) $ICC_{2,1}$ = 0.88 (0.73-0.95) $ICC_{2,1}$ = 0.96 (0.91-0.98) $ICC_{2,1}$ = 0.96 (0.91-0.98)	2.47 3.93 3.08 3.16 2.43 4.15 3.06 2.82	6.6 9.8 6.6 6.4 6.9 11.0 6.7 6.0	6.85 (1) 10.91 (1) 8.54 (1) 8.77 (1) 6.74 (1) 11.51 (1) 8.48 (1) 7.81 (1)	18.4 (1) 27.1 (1) 18.2 (1) 17.6 (1) 19.1 (1) 30.4 (1) 18.7 (1) 16.6 (1)

(Continues)

Author/year	Movement	Type of reliability	Test condition	Reliability: ICC_{model} (95% CI)/*r*	Measurement error: SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
	ER	Intra	Highest value,	$ICC_{3,1} = 0.94$	2.28	8.5	6.32 (1)	23.5 (1)
			EFD, left	(0.86-0.98)	2.21	8.4	6.12 (1)	23.4 (1)
			Highest value,	$ICC_{3,1} = 0.93$	3.20	8.4	8.86 (1)	23.2 (1)
			EFD, right	(0.83-0.97)	1.96	4.8	5.43 (1)	13.3 (1)
			Highest value,	$ICC_{3,1} = 0.92$	2.02	8.0	5.61 (1)	22.1 (1)
			HHD, left	(0.81-0.97)	1.78	7.2	4.94 (1)	19.9 (1)
			Highest value,	$ICC_{3,1} = 0.97$	2.59	7.1	7.19 (1)	19.7 (1)
			HHD, right	(0.93-0.99)	2.74	7.1	7.58 (1)	19.5 (1)
			Average value, EFD, left	$ICC_{3,1} = 0.95$ (0.87-0.99)				
			Average value, EFD, right	$ICC_{3,1} = 0.95$ (0.88-0.98)				
			Average value, HHD, left	$ICC_{3,1} = 0.94$ (0.85-0.98)				
			Average value, HHD, right	$ICC_{3,1} = 0.94$ (0.86-0.98)				
		Inter	Highest value,	$ICC_{2,1} = 0.94$	2.39	8.9	6.62 (1)	24.6 (1)
			EFD, left	(0.84-0.97)	2.20	8.4	6.11 (1)	23.3 (1)
			Highest value,	$ICC_{2,1} = 0.93$	3.27	8.6	9.06 (1)	23.7 (1)
			EFD, right	(0.84-0.97)	2.32	5.7	6.44 (1)	15.8 (1)
			Highest value,	$ICC_{2,1} = 0.92$	2.07	8.1	5.74 (1)	22.6 (1)
			HHD, left	(0.80-0.97)	1.73	7.0	4.79 (1)	19.3 (1)
			Highest value,	$ICC_{2,1} = 0.96$	2.70	7.4	7.49 (1)	20.5 (1)
			HHD, right	(0.85-0.99)	3.23	8.3	8.96 (1)	23.1 (1)
			Average value, EFD, left	$ICC_{2,1} = 0.95$ (0.87-0.97)				
			Average value, EFD, right	$ICC_{2,1} = 0.95$ (0.89-0.98)				
			Average value, HHD, left	$ICC_{2,1} = 0.94$ (0.85-0.97)				
			Average value, HHD, right	$ICC_{2,1} = 0.92$ (0.69-0.97)				
Johansson, 2005	Abd	Intra	Pull force, tester A, mean	$ICC_{?} = 0.98$ (0.95-0.99)				
			Pull force, tester A, max	$ICC_{?} = 0.98$ (0.96-0.99)				
			Resisted force, tester A, mean	$ICC_{?} = 0.94$ (0.88-0.98)				
			Resisted force, tester A, max	$ICC_{?} = 0.95$ (0.78-0.98)				
			Pull force, tester B, mean	$ICC_{?} = 0.96$ (0.91-0.98)				
			Pull force, tester B, max	$ICC_{?} = 0.95$ (0.91-0.98)				
			Resisted force, tester B, mean	$ICC_{?} = 0.90$ (0.78-0.95)				
			Resisted force, tester B, max	$ICC_{?} = 0.91$ (0.80-0.96)				

(Continues)

Author/year	Movement	Type of reliability	Test condition	Reliability ICC_{model} (95% CI)/*r*	Measurement error SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
	Abd	Inter	Pull force, test 1, mean Pull force, test 1, max Resisted force, test 1, mean Resisted force, test 1, max Pull force, test 2, mean Pull force, test 2, max Resisted force, test 2, mean Resisted force, test 2, max	$ICC_{?}$ = 0.93 (0.87-0.96) $ICC_{?}$ = 0.96 (0.91-0.98) $ICC_{?}$ = 0.89 (0.78-0.94) $ICC_{?}$ = 0.91 (0.82-0.96) $ICC_{?}$ = 0.89 (0.80-0.94) $ICC_{?}$ = 0.97 (0.94-0.98) $ICC_{?}$ = 0.94 (0.79-0.98) $ICC_{?}$ = 0.95 (0.82-0.98)				
Johansson, 2015	ER	Intra	Examiner 1 Examiner 2	$ICC_{2,1}$ = 0.88 (0.78-0.94) $ICC_{2,1}$ = 0.86 (0.73-0.93)	11.10 7.85	9.2 6.8		25.5 (c) 18.8 (c)
		Inter		$ICC_{2,1}$ = 0.71 (0.42-0.87)	14.11	12.0		33.3 (c)
Leggin, 1996	IR	Intra	Nicholas, examiner A Nicholas, examiner B	$ICC_{3,1}$ = 0.94 $ICC_{3,1}$ = 0.97				
		Inter	Nicholas	$ICC_{3,1}$ = 0.90				
	ER	Intra	Nicholas, examiner A Nicholas, examiner B	$ICC_{3,1}$ = 0.95 $ICC_{3,1}$ = 0.89				
		Inter	Nicholas	$ICC_{3,1}$ = 0.94				
	Abd	Intra	Nicholas, examiner A Nicholas, examiner B Isomex 2.0, examiner A Isomex 2.0, examiner B	$ICC_{3,1}$ = 0.84 $ICC_{3,1}$ = 0.96 $ICC_{3,1}$ = 0.98 $ICC_{3,1}$ = 0.98				
		Inter	Nicholas Isomex 2.0	$ICC_{3,1}$ = 0.79 $ICC_{3,1}$ = 0.92				
Magnusson, 1990	Abd	Intra	Day 1 vs. day 2 Day 1 vs. day 3 Day 1 vs. day 4 Day 1 vs. day 5	*r* = .98 *r* = .94 *r* = .96 *r* = .96				

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability ICC_{model} (95% CI)/*r*	Measurement error SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = %MDC_{95} (c) = %MDC_{95} calculated
McLaine, 2016	IR	Intra	Dom., sitting Dom., supine Dom., prone Non-dom., sitting Non-dom., supine Non-dom., prone	$ICC_{3,1}$ = 0.97 (0.90-0.96) $ICC_{3,1}$ = 0.93 (0.80-0.98) $ICC_{3,1}$ = 0.97 (0.90-0.99) $ICC_{3,1}$ = 0.90 (0.70-0.97) $ICC_{3,1}$ = 0.97 (0.91-0.99) $ICC_{3,1}$ = 0.94 (0.82-0.98)	2.41 4.90 2.89 5.75 2.13 4.96		5.63 (2) 11.42 (2) 6.74 (2) 13.41 (2) 4.97 (2) 11.57 (2)	6.4 (c)) 13.6 (c) 3.4 (c) 16.0 (c) 5.8 (c) 13.7 (c)
	ER	Intra	Dom., sitting Dom., supine Dom., prone Non-dom., sitting Non-dom., supine Non-dom., prone	$ICC_{3,1}$ = 0.97 (0.88-0.99) $ICC_{3,1}$ = 0.96 (0.87-0.99) $ICC_{3,1}$ = 0.95 (0.81-0.99) $ICC_{3,1}$ = 0.97 (0.91-0.99) $ICC_{3,1}$ = 0.92 (0.68-0.97) $ICC_{3,1}$ = 0.96 (0.89-0.99)	1.79 3.15 3.31 2.36 4.97 2.99		4.17 (2) 7.36 (2) 7.73 (2) 5.51 (2) 11.6 (2) 6.98 (2)	5.5 (c) 8.2 (c) 9.1 (c) 7.1 (c) 12.4 (c) 38.3 (c)
	Flex	Intra	Dom., sitting Dom., supine Dom., prone Non-dom., sitting Non-dom., supine Non-dom., prone	$ICC_{3,1}$ = 0.94 (0.82-0.98) $ICC_{3,1}$ = 0.94 (0.82-0.98) $ICC_{3,1}$ = 0.87 (0.62-0.96) $ICC_{3,1}$ = 0.93 (0.78-0.98) $ICC_{3,1}$ = 0.94 (0.81-0.98) $ICC_{3,1}$ = 0.93 (0.79-0.98)	2.15 2.76 2.38 2.70 2.67 1.51		5.02 (2) 6.43 (2) 5.54 (2) 6.30 (2) 6.21 (2) 3.53 (2)	12.6 (c) 12.2 (c) 18.2 (c) 16.5 (c) 12.1 (c) 11.8 (c)
	Ex	Intra	Dom., sitting Dom., supine Dom., prone Non-dom., sitting Non-dom., supine Non-dom., prone	$ICC_{3,1}$ = 0.96 (0.88-0.99) $ICC_{3,1}$ = 0.98 (0.94-0.99) $ICC_{3,1}$ = 0.98 (0.95-0.99) $ICC_{3,1}$ = 0.97 (0.90-0.99) $ICC_{3,1}$ = 0.96 (0.87-0.99) $ICC_{3,1}$ = 0.99 (0.97-0.99)	1.86 1.74 1.51 1.91 3.53 0.78		4.34 (2) 4.05 (2) 3.52 (2) 4.45 (2) 8.25 (2) 1.82 (2)	8.6 (c) 6.7 (c) 5.2 (c) 8.7 (c) 13.1 (c) 2.7 (c)
McMahon, 1992	Abd	Intra	WS BS	$ICC_{?}$ = 0.88 $ICC_{?}$ = 0.83				

(Continues)

Appendix 2 (Continued)

				Reliability	Measurement error			
Author/year	Movement	Type of reliability	Test condition	ICC_{model} (95% CI)/r	SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
		Inter	WS	$ICC_{?} = 0.81$				
O'Shea, 2007	Flex	Intra		$ICC_{2,1} = 0.81$				
Ottenbacher, 2002	Abd	Inter	Part 1, 0° abd Part 1, 90 ° abd Part 2, 0° abd Part 2, 90 ° abd	$ICC_{3,1} = 0.90$ $ICC_{3,1} = 0.79$ $ICC_{3,1} = 0.98$ $ICC_{3,1} = 0.93$				
Phillips, 2000	ER	Intra	WS, dom. WS, non-dom. BS, dom. BS, non-dom.	$ICC_{1,3} = 0.99$ (0.97-0.99) $ICC_{1,3} = 0.99$ (0.98-1.00) $ICC_{1,1} = 0.95$ (0.87-0.97) $ICC_{1,1} = 0.98$ (0.93-0.99)				
	Abd	Intra	WS, dom. WS, non-dom. BS, dom. BS, non-dom.	$ICC_{1,3} = 0.97$ (0.94-0.99) $ICC_{1,3} = 0.98$ (0.96-0.99) $ICC_{1,1} = 0.89$ (0.71-0.93) $ICC_{1,1} = 0.90$ (0.73-0.93)				
Riemann, 2010	IR	Intra	Left, tester 1, prone, 90° Left, tester 1, seated, neutral Left, tester 1, seated, 30° Left, tester 2, prone, 90° Left, tester 2, seated, neutral Left, tester 2, seated, 30° Right, tester 1, prone, 90° Right, tester 1, seated, neutral Right, tester 1, seated, 30° Right, tester 2, prone, 90° Right, tester 2, seated, neutral Right, tester 2, seated, 30°	$ICC_{3,1} = 0.88$ $ICC_{3,1} = 0.87$ $ICC_{3,1} = 0.86$ $ICC_{3,1} = 0.84$ $ICC_{3,1} = 0.69$ $ICC_{3,1} = 0.76$ $ICC_{3,1} = 0.89$ $ICC_{3,1} = 0.80$ $ICC_{3,1} = 0.84$ $ICC_{3,1} = 0.81$ $ICC_{3,1} = 0.74$ $ICC_{3,1} = 0.57$				

(Continues)

Appendix 2 (Continued)

				Reliability	Measurement error			
Author/year	**Movement**	**Type of reliability**	**Test condition**	**ICC$_{model}$ (95% CI)/*r***	**SEM, N**	**%SEM**	**(1) = MDC$_{95}$, N (2) = MDC$_{90}$, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC$_{95}$ calculated**	**(1) = %MDC$_{95}$ (c) = %MDC$_{95}$ calculated**
		Inter	Left, prone, 90° Left, seated, neutral Left, seated, 30° Right, prone, 90° Right, seated, neutral Right, seated, 30°	ICC$_{3,1}$ = 0.72 ICC$_{3,1}$ = 0.90 ICC$_{3,1}$ = 0.90 ICC$_{3,1}$ = 0.87 ICC$_{3,1}$ = 0.75 ICC$_{3,1}$ = 0.81				
	ER	Intra	Left, tester 1, prone, 90° Left, tester 1, seated, neutral Left, tester 1, seated, 30° Left, tester 2, prone, 90° Left, tester 2, seated, neutral Left, tester 2, seated, 30° Right, tester 1, prone, 90° Right, tester 1, seated, neutral Right, tester 1, seated, 30° Right, tester 2, prone, 90° Right, tester 2, seated, neutral Right, tester 2, seated, 30°	ICC$_{3,1}$ = 0.85 ICC$_{3,1}$ = 0.70 ICC$_{3,1}$ = 0.92 ICC$_{3,1}$ = 0.77 ICC$_{3,1}$ = 0.77 ICC$_{3,1}$ = 0.87 ICC$_{3,1}$ = 0.79 ICC$_{3,1}$ = 0.94 ICC$_{3,1}$ = 0.88 ICC$_{3,1}$ = 0.80 ICC$_{3,1}$ = 0.80 ICC$_{3,1}$ = 0.77				
		Inter	Left, prone, 90° Left, seated, neutral Left, seated, 30° Right, prone, 90° Right, seated, neutral Right, seated, 30°	ICC$_{3,1}$ = 0.64 ICC$_{3,1}$ = 0.78 ICC$_{3,1}$ = 0.85 ICC$_{3,1}$ = 0.82 ICC$_{3,1}$ = 0.76 ICC$_{3,1}$ = 0.92				
Sciascia, 2015	Flex		Asymptomatic, dom. Asymptomatic, non-dom. Symptomatic, dom. Symptomatic, non-dom.	ICC$_{2,1}$ = 0.98 (0.95-0.99) ICC$_{2,1}$ = 0.98 (0.95-0.99) ICC$_{2,1}$ = 0.97 (0.91-0.99) ICC$_{2,1}$ = 0.95 (0.86-0.98)			9.81 (2) 9.81 (2) 19.62 (2) 19.62 (2)	9.6 (c) 9.6 (c) 17.7 (c) 17.7 (c)

(Continues)

Appendix 2 (Continued)

Author/year	Movement	Type of reliability	Test condition	Reliability: ICC_{model} (95% CI)/*r*	Measurement error: SEM, N	%SEM	(1) = MDC_{95}, N (2) = MDC_{90}, N (3) = TEM, N (4) = LoA, kg (5) = LoA, N (6) = SDD, N (c) = MDC_{95} calculated	(1) = $\%MDC_{95}$ (c) = $\%MDC_{95}$ calculated
Sullivan, 1988	ER	Intra		*r* = .99				
Vermeulen, 2005	ER	Intra	HHD EFD	$ICC_{2,1}$ = 0.91 (0.80-0.97) $ICC_{2,1}$ = 0.90 (0.77-0.96)			26.5 (6) 31.5 (6)	
		Inter	HHD EFD	$ICC_{2,1}$ = 0.88 (0.71-0.95) $ICC_{2,1}$ = 0.86 (0.68-0.94)			30.8 (6) 36.6 (6)	
	Abd	Intra	HHD EFD	$ICC_{2,1}$ = 0.86 (0.68-0.94) $ICC_{2,1}$ = 0.82 (0.61-0.93)			53.5 (6) 78.2 (6)	
		Inter	HHD EFD	$ICC_{2,1}$ = 0.89 (0.74-0.96) $ICC_{2,1}$ = 0.90 (0.77-0.96)			39.6 (6) 57.2 (6)	
	Flex	Intra	HHD EFD	$ICC_{2,1}$ = 0.86 (0.68-0.94) $ICC_{2,1}$ = 0.88 (0.71-0.95)			31.5 (6) 24.3 (6)	
		Inter	HHD EFD	$ICC_{2,1}$ = 0.86 (0.68-0.94) $ICC_{2,1}$ = 0.90 (0.76-0.96)			29.6 (6) 20.8 (6)	

Abbreviations: Abd, abduction; Add, adduction; BS, between sessions; dom., dominant; EFD, externally fixed dynamometer; ER, external rotation; EX, extension; Flex, flexion; HHD, handheld dynamometer; ICC, intraclass correlation coefficient; Inter, inter-tester reliability; Intra, intra-tester reliability; IR, internal rotation; kg, kilogram; LoA, limits of agreement; MDC, minimal detectable change; MDC90, MDC with 90% CI; MDC95, MDC with 95% CI; N, Newton; non-dom., non-dominant; r, Pearson's correlation coefficient; SDD, smallest detectable difference; SEM, standard error of measurement; SIS, subacromial impingement syndrome; Stab., stabilization; TEM, technical error of the measurement; WS, within session.

Appendix 3

Results of construct validity

Author/year	Comparison	Movement	Test procedure	Results
Burnham, 1995	HHD/ ID	Abd	Isometric	$r = .28$
			60°/s	$r = .43$
Hayes, 2002	HHD/ spring scale	IR		$r = .82$
		ER		$r = .77$
		flex		$r = .88$
Holt, 2016	HHD/ ID	IR	Left, ecc, 60°/s	$r = .75$
			Left, ecc, 180°/s	$r = .77$
			Left, ecc, 240°/s	$r = .77$
			Left, con, 60°/s	$r = .85$
			Left, con, 180°/s	$r = .80$
			Left, con, 240°/s	$r = .79$
			Right, ecc, 60°/s	$r = .57$
			Right, ecc, 180°/s	$r = .71$
			Right, ecc, 240°/s	$r = .65$
			Right, con, 60°/s	$r = .66$
			Right, con, 180°/s	$r = .70$
			Right, con, 240°/s	$r = .76$
		ER	Left, ecc, 60°/s	$r = .77$
			Left, ecc, 180°/s	$r = .75$
			Left, ecc, 240°/s	$r = .45$
			Left, con, 60°/s	$r = .62$
			Left, con, 180°/s	$r = .63$
			Left, con, 240°/s	$r = .65$
			Right, ecc, 60°/s	$r = .78$
			Right, ecc, 180°/s	$r = .70$
			Right, ecc, 240°/s	$r = .70$
			Right, con, 60°/s	$r = .85$
			Right, con, 180°/s	$r = .77$
			Right, con, 240°/s	$r = .79$
Johansson, 2005	HHD/ spring balance	Abd	Pull force, mean	$ICC_{?} = 0.96$ (0.91-0.99)
			Pull force, max	$ICC_{?} = 0.96$ (0.91-0.99)
			Resisted force, mean	$ICC_{?} = 0.99$ (0.98-1.00)
			Resisted force, max	$ICC_{?} = 0.99$ (0.96-0.99)
Sullivan, 1998	HHD/ ID	ER		$r = .52$
				$r = .78$
Vermeulen, 2005	HHD/ EFD	ER		Mean diff, N = 21.4 (15.9-26.8)
		Abd		Mean diff, N = −6.5 (−16.3-3.3)
		Flex		Mean diff, N = 29.9 (26.3-33.5)
Sciascia, 2016	Subjects with/ without shoulder symptoms	Flex		*P*-value = .89
				P-value = .99

Abbreviations: Abd, abduction; con, concentric, Pearson's correlation coefficient; diff, difference; ecc, eccentric; EFD, externally fixed dynamometer; ER, external rotation; Flex, flexion; HHD, handheld dynamometer; ID, isometric dynamometer; IR, internal rotation.

P-value: independent t test