Measurement properties of isokinetic dynamometry for assessment of shoulder muscle strength

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

published in
Archives of Physical Medicine and Rehabilitation
2021

DOI (link to publisher)
10.1016/j.apmr.2020.06.005

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2021;102:510-20

SYSTEMATIC REVIEW

Measurement Properties of Isokinetic Dynamometry for Assessment of Shoulder Muscle Strength: A Systematic Review

Lotte Sørensen, MSc,[a,b] Lisa Gregersen Oestergaard, PhD,[a,c] Maurits van Tulder, PhD,[a,d] Annemette Krintel Petersen, PhD[a,b]

From the [a]Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Aarhus, Denmark; [b]Department of Clinical Medicine, Aarhus University, Aarhus, Denmark; [c]DEFACTUM, Central Denmark Region, Aarhus, Denmark; and [d]Department of Health Sciences, Amsterdam Movement Sciences Research Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

Abstract

Objective: To investigate the evidence of measurement properties of isokinetic dynamometry (ID) for assessment of shoulder muscle strength in healthy individuals and patients with nonneurologic shoulder pathology.

Data Sources: Cochrane Central Register of Controlled Trials, PubMed, EMBASE, and Physiotherapy Evidence Database were searched up to February 2020 without restrictions. Reference lists and citations were hand-searched.

Study Selection: Two review authors independently included studies that met the following criteria: (1) evaluated measurement properties of ID when used on the glenohumeral joint and (2) included individuals 18 years and older. Studies including patients with neurologic, neuromuscular, or systemic diseases or critical illness were excluded.

Data Extraction: The quality assessment and data synthesis were performed according to the COnsensus-based Standards for the selection of health Measurement INstruments methodology.

Data Synthesis: Twenty-one studies with a total of 597 participants were included. The results were combined separately for isometric, concentric, and eccentric test mode; for the velocities 30°/s-60°/s, 90°/s, 120°/s, and 240°/s; for the seated, supine, and standing position; and for internal rotation (IR), external rotation (ER), and the ER/IR ratio. The reliability of ID was overall sufficient with the majority of intraclass correlation coefficients $\geq$0.70. The quality of evidence was moderate or low for 20 of 30 strata examined. The measurement error results were rated as insufficient for all strata. The SEM ranged from 4%-28%. The quality of evidence varied depending of strata examined.

Conclusions: The reliability of ID for measurement of shoulder strength was overall sufficient for all positions, velocities, and modes of strength. The measurement error was not sufficient. Because most studies used the seated position, the velocities 30°/s-60°/s or 120°/s, and the concentric test mode, the quality of evidence was highest for these conditions.

Archives of Physical Medicine and Rehabilitation 2021;102:510-20

Assessment of shoulder muscle strength is frequently used in both athletes and patients with shoulder disorders.[1,2] Strength evaluation of the shoulder muscles with dynamometers is a useful method for clinicians to objectively assess muscle strength, quantify the degree of impairment, guide treatment, and evaluate treatment efficacy.[2-4] Furthermore, muscle strength measurement is used to prevent injuries and improve performance.

Isokinetic dynamometry (ID) is a commonly used objective method for measuring muscle strength.[1,5] ID can measure muscle strength with accommodating resistance at a constant angular velocity and assess the maximal torque production throughout a prescribed range of motion.[2,6] Isokinetic dynamometers can evaluate muscle strength in isometric, concentric, or eccentric test mode and across a wide range of speeds.[3] Like any assessment methodology, ID must be valid and reliable to be meaningful in

The study was funded by Aarhus University, Denmark and Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Denmark.

Clinical Trial Registration No.: CRD42017054027.

Disclosures: none.

0003-9993/20/$36 - see front matter © 2020 by the American Congress of Rehabilitation Medicine

https://doi.org/10.1016/j.apmr.2020.06.005

clinical practice and research.[7,8] The more reliable the measurements, the better precision of single measurements and better tracking of changes.[7] Reliability refers to the consistency of a measurement method and reflects both reliability (relative reliability) and measurement error (absolute reliability).[9] Reliability refers to the consistency of a test or measurement and is defined as a ratio of the variation between individuals divided by the total variation. The total variation consists of variation between individuals and measurement error.[8,9] Reliability assesses whether subjects can be distinguished from each other despite measurement errors and is highly dependent on the heterogeneity of the study population.[7-9] Measurement error quantifies the systematic and random error of a score that is not attributed to true changes in the construct to be measured. It assesses how close to each other the results of repeated measurements are and is expressed in the unit of the measurement.[8-10]

Few systematic reviews have summarized the results of some of the issues related to the assessment of shoulder muscle strength with ID. Edouard et al[3] conducted a systematic review to determine the influence of position on the reliability of internal rotation (IR) and external rotation (ER) isokinetic strength assessment. Rabelo et al[2] investigated the reliability of muscle strength assessment of 5 different joints in chronic poststroke hemiparesis.

In general, clinicians are interested in information about the most reliable protocol for a given test procedure in a specific population. Therefore, the aims of this systematic review were to (1) summarize the evidence of measurement properties of isokinetic dynamometry for the assessment of shoulder muscle strength in healthy individuals and patients with nonneurologic shoulder pathology and (2) to evaluate if population, test mode, position, and velocity affect the measurement properties of isokinetic dynamometry.

Methods

We performed this systematic review in accordance with the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) methodology for systematic reviews of Patient-Reported Outcome Measures.[11] When needed, we adapted COSMIN to the purpose of evaluating the quality of objective measurement instruments. The COSMIN methodology is based on existing guidelines for reviews, such as the Preferred Reporting Items for Systematic Reviews and Meta-Analysis statement,[12] the Cochrane Handbook for systematic reviews of interventions,[13] and the Grading of Recommendations Assessment, Development, and Evaluation principles.[14]

Two review authors (A.K.P., L.S.) independently conducted the study selection, data extraction, quality assessment, and results rating. Disagreements were resolved through discussion, and when needed, a third review author (L.G.O.) was consulted.

List of abbreviations:

COSMIN	**COnsensus-based Standards for the selection of health Measurement INstruments**
ER	**external rotation**
ICC	**intraclass correlation coefficient**
ID	**isokinetic dynamometry**
IR	**internal rotation**
MDC	**minimal detectable change**
MIC	**minimal important change**

Deviation from the protocol

Before starting the review, a protocol was registered in the International Prospective Register of Systematic Reviews (registration no. CRD42017054027) and published.[15] The original purpose of the review as described in the protocol was to summarize the evidence of both ID and hand-held dynamometry. We intended to report the results separately but in 1 review. After full-text assessment of eligibility we considered the data to be too comprehensive for a single review and decided to divide the review into 2 separate reviews; one review focusing on ID and another on hand-held dynamometry. The current review is focusing on ID.

Search strategy

We searched several electronic databases from their inception to February 2020: Cochrane Central Register of Controlled Trials, PubMed, EMBASE, and Physiotherapy Evidence Database. An information specialist helped us developing the electronic search strategies, using MeSH (PubMed), Thesaurus (EMBASE), and free text words. We combined these terms with the validated sensitive methodological PubMed search filter developed to identify studies on measurement properties of measurement instruments[16] and an adapted version of this filter for searching EMBASE. The search strategy for PubMed is presented in supplemental appendix S1 (available online only at http://www.archives-pmr.org/). We hand-searched reference lists and citations of the included studies and relevant reviews for additional studies. We did not apply any restrictions on publication period or language. We included only studies published as full-text original articles.

Study selection

To be included in this review, studies needed to (1) have evaluated measurement properties of ID when used on the glenohumeral joint and (2) have included individuals 18 years and older with or without shoulder symptoms. Exclusion criteria were studies that included patients with neurologic, neuromuscular, or systemic diseases or critical illness.

We screened title and abstracts and obtained full text of all potentially relevant studies to identify studies meeting the inclusion criteria. We recorded any reasons for exclusion of retrieved full-text articles (data not shown).

Data extraction

We extracted data on characteristics of the instrument and included study population (age, sex, healthy/symptomatic individuals), test procedure, and results of the measurement properties.

Quality assessment

We used the COSMIN Risk of Bias checklist to assess the methodological quality of each study.[11,17] The checklist consists of 10 boxes that contain items about design aspects and statistical methods. We evaluated the measurement properties examined in each article using the corresponding COSMIN box. We rated the methodological quality of each item in a box as "very good," "adequate," "doubtful," or "inadequate." We used the lowest rating in the box for determining the overall quality of a measurement property in each study ("the worst score counts" principle).

Outcomes

Following COSMIN, we rated the results of each study as either sufficient (+), insufficient (−), or indeterminate (?). We based this rating on the criteria for good measurement properties.[11,17,18] We rated reliability as sufficient if intraclass correlation coefficient (ICC) was ≥0.70 and measurement error as sufficient if minimal detectable change (MDC) was less than or equal to minimal important change (MIC).[11] Consensus on a definition of MIC for muscle strength is not available, but some studies suggest a change of 10%-15% as clinically relevant.[4,19] Based on available literature and clinical reasoning, we set the MIC to 15%. We conducted a sensitivity analysis with the MIC being set at 10% and 20%, respectively.

In studies included in this review, SEM was a more consistent estimate for measurement error than MDC. Therefore, instead of rating the %MDC results against the MIC criteria, we chose to rate the %SEM. Based on the formula $MDC = SEM \times 1.96 \times \sqrt{2}$ the %SEM equivalent to 15% MDC is 5.4%. Therefore, we rated if %SEM was ≤5.4%.

Some studies only reported SEM as absolute peak torque values expressed in newton meters. If the study presented sufficient data by reporting the actual muscle strength results, we calculated the %SEM by dividing SEM with the mean of test and retest.

Evidence synthesis

When results from different studies were consistent, we summarized these results to determine the overall evidence of the measurement properties of ID.[11] We combined results separately in strata for isometric, concentric (con) and eccentric (ecc) test mode; for the velocities 30°/s-60°/s, 90°/s, 120°/s, and 240°/s; for the seated, supine, and standing position; and for IR, ER, and the ER/IR ratio. We present the results this way for clinicians, who are generally interested in a specific movement, test mode, position, and velocity.

For each stratum, we reported the ICC results as a range of minimal and maximal values and as the proportion of all the reported ICC values ≥0.70. The %SEM results were reported as a range between minimal and maximal values and as the proportion of reported %SEM estimates ≤5.4%.

Based on COSMIN, we rated the summarized results for each stratification against the criteria for good measurement properties, to determine if the overall measurement property was sufficient (+), insufficient (−), inconsistent (±), or indeterminate (?). We rated "sufficient" if ≥75% of the ICC or %SEM values met the criteria, "insufficient" if ≤25% of the ICC or %SEM values met the criteria, and "indeterminate" if between 25% and 75% of the ICC or %SEM values met the criteria.[11]

Furthermore, we performed a sensitivity analysis reporting the proportion of %SEM estimates ≤3.6% (equivalent to %MDC≤10%) and %SEM estimates ≤7.2% (equivalent to %MDC≤20%).

We used the Grading of Recommendations Assessment, Development, and Evaluation approach, modified for reviews of measurement properties, to classify the quality of evidence as "high," "moderate," "low," or "very low."[11,20] We used risk of bias, inconsistency, imprecision, and indirectness to downgrade the evidence.[11] Risk of bias refers to the methodological quality of the studies, inconsistency to inconsistency of the results of the studies within the pooled stratifications, imprecision to the total sample size of the included studies (downgraded with 1 level if the sample size was below 100 and downgraded with 2 levels if the sample size was below 50), and indirectness to the circumstance where included studies were partly performed in another population or context.[11]

Results

The electronic search strategy yielded 8054 hits, and 6368 remained after removal of duplicates. We excluded 6267 studies based on title and abstract, leaving 101 studies for full-text assessment. We excluded another 84 studies after evaluating full text. We identified 4 additional studies through reference checking. In total, we included 21 studies in this review (fig 1).

Characteristics of the included studies are shown in table 1. The study populations included 563 healthy individuals in 19 studies and 34 individuals with shoulder pathology in 2 studies. Supplemental appendix S2 (available online only at http://www.archives-pmr.org/) shows the measurement property results extracted for each single study. Rating of the results and risk of bias assessment are presented in table 2.

Results were reported separately for isometric, concentric, and eccentric test mode; for the velocities 30°/s-60°/s, 90°/s, 120°/s, and 240°/s; for the seated, supine, and standing position; and for the movements IR and ER. Few studies evaluated abduction, adduction, flexion, and extension[23,30-32,35,36]; these results are reported in supplemental appendix S2 but not included in the summarized results. Results of ER_{con}/IR_{con} ratio for the velocities 60°/s and 120°/s in the seated position were summarized. As seen from supplemental appendix S2, 1 study[24] evaluated other ratios (ER_{ecc}/IR_{con}, ER_{ecc}/IR_{ecc}, IR_{ecc}/ER_{con}); these results are not included in the summarized results. The most commonly used unit to assess muscle strength was peak torque. Only 2 studies[22,33] reported total work and average power in addition to peak torque; we do not present these results in the text.

Reliability

All 21 studies evaluated reliability in a test-retest design.[6,21-40] We present the summarized results, the overall rating of the results, and the quality of evidence in the summary of findings table (table 3). Most studies reported the ICC, but 5 studies used Pearson correlation coefficient, variability, or coefficient of variation.[25,29,32,35,40] We did not include these results in the text, but they are reported in supplemental appendix S2. Eleven studies reported ICC but without 95% CI.[21,26,28,30,31,33,34,36-39]

Of the strata examined, 28 of 30 showed at least 75% of the ICC results ≥0.70 and the overall rating of the results was "sufficient." The IR_{con} at 90°/s revealed 66% of the results ≥0.70, and the ER_{con}/IR_{con} ratio at 120°/s revealed 50% of the results ≥0.70; these strata were rated as "indeterminate" (see table 3). The quality of evidence was "high" in 2 strata, "moderate" in 10 strata, "low" in 10 strata, and "very low" in 8 strata (see table 3).

Two studies examined the reliability of ID in a total of 34 patients with shoulder pathology and compared affected and nonaffected shoulder. Both studies[21,33] found higher reliability for the affected shoulder than for the nonaffected shoulder.

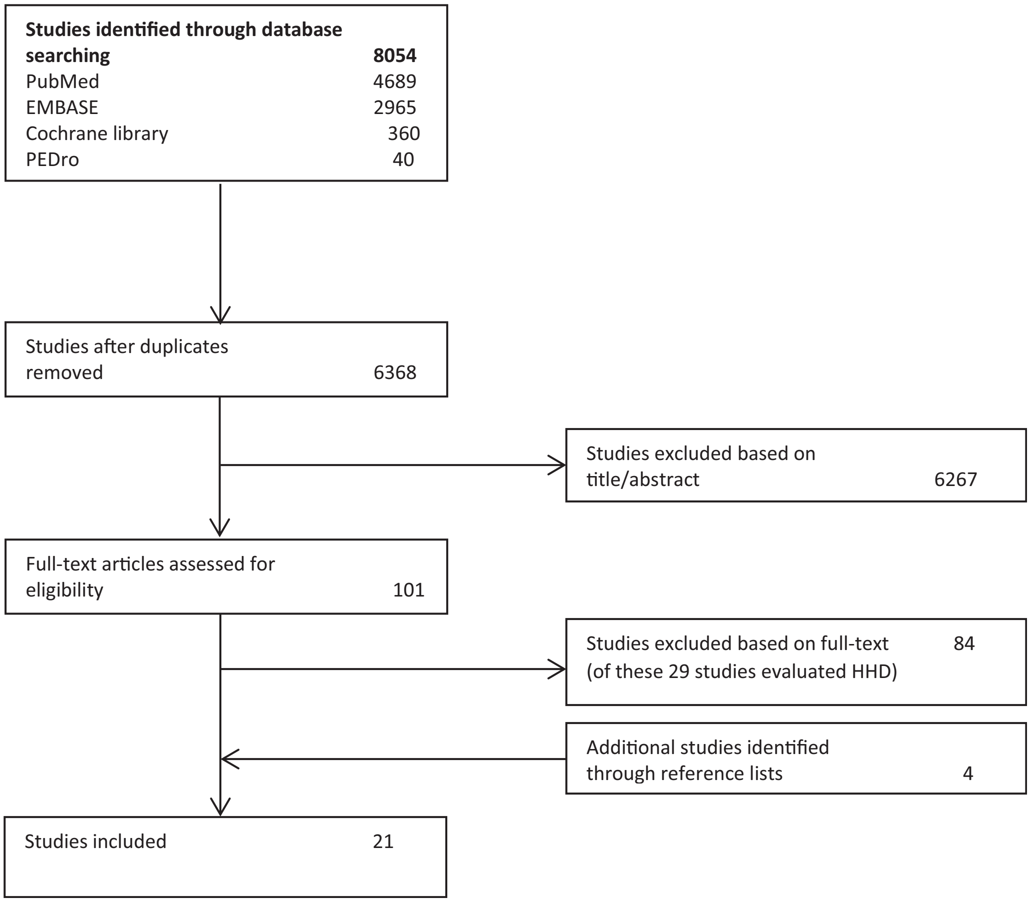

Fig 1 Flowchart of studies through the selection process. Abbreviations: HHD, hand-held dynamometry; PEDro, Physiotherapy Evidence Database.

Seven studies evaluated the reliability of eccentric as well as concentric test mode.[6,22,24,29,33,34,38] Of these, 2 studies[6,34] found the reliability of the eccentric test mode to be lower than of the concentric test mode. The other 5 studies found no clear difference between test modes.

One study[28] examined the seated and supine position. No conclusion could be made about the optimal test position. None of the velocities investigated clearly showed higher reliability than others.

Measurement error

Measurement error was investigated in 13 studies.[6,21-25,27,29,31,34,36-38] The summarized results, the overall rating of the results, and the quality of evidence are presented in the Summary of Findings table (table 4). Seven studies reported SEM only as an absolute value but provided sufficient data to calculate the %SEM.[21,23,25,29,34,37,38] One study reported SEM in pounds, which were converted into newtons (1lb=4.448N).[27] All results calculated by the review team are clearly stated in supplemental appendix S2.

The %SEM values ranged from 4.4%-27.8% (equivalent to %MDC values from 12%-77%), and all strata examined were rated as “insufficient.” The quality of evidence was “high” in 2 strata, “moderate” in 6 strata, “low” in 5 strata, and “very low” in 9 strata (see table 4).

One study of 10 patients found larger measurement error for the affected shoulder than for the nonaffected shoulder.[21]

Seven studies examined the measurement error of eccentric as well as concentric test mode.[6,22,24,29,34,35,38] Three of these studies found a tendency toward larger measurement error for the eccentric test mode than for the concentric test mode.[6,24,35] The other 4 studies found no clear difference between test modes. One study found smaller measurement error for the supine position than for the seated position.[25]

None of the velocities investigated clearly showed smaller measurement error than others.

Because all the results were rated as “insufficient” when MIC was set to 15%, we did not perform the sensitivity analysis with MIC at 10%. By increasing the MIC to 20% (equivalent to %SEM$\leq$7.2) the results of all strata were still rated as “insufficient.”

Discussion

The results of this systematic review demonstrated that the reliability of ID was overall sufficient, but the quality of evidence was mostly “moderate” or “low.” We rated the measurement error results as “insufficient” for all strata investigated. This rating was consistent when MDC was rated to be less or equal to 15% and 20%, respectively. The quality of evidence was mostly “moderate,” “low,” or “very low.”

Table 1 Characteristics of included studies

Author	Participants	Type of Dynamometer	Procedure	Movement/Velocity	Unit Measurement
Anderson et al[21]	N=10 Condition: Subjects with chronic rotator cuff pathology Age: 62.4±9.5 y Sex: 6 male, 4 female	KinCom	Time test-retest: 2 d Position: Seated Bilateral 1 examiner, physiotherapist	IR+ER Concentric 60°/s	Peak torque (N•m) 3 repetitions Highest value
Dauty et al[22]	N=14 Condition: Healthy subjects Age: 47.5±5.6 y Sex: 7 male, 7 female	Cybex	Time test-retest: 26±4 d Position: Seated Bilateral 1 examiner	ER: concentric+eccentric IR: concentric 60°/s+120°/s $ER_{ecc}/IR_{con}+ER_{con}/IR_{con}$	Peak torque (N•m) Work (J) Power (W) 5 repetitions Highest or average value?
Cavuoto et al[23]	N=142 Condition: Healthy subjects Age: Young adults Sex: 71 male, 71 female	Cybex	Time test-retest: at least 48 h Position: Supine Right shoulder Examiners not described	Flex Isometric	Peak torque (N•m) 3 repetitions in 4 sessions Highest value
Edouard et al[24]	N=46 Condition: Healthy subjects Age: 38±13 y Sex: 24 male, 22 female	Biodex	Time test-retest:7 d Position: Seated Bilateral 1 examiner	IR+ER Concentric: 60°/s+120°/s Eccentric: 30°/s Different ER/IR+IR/IR+ER/ER ratios	Peak torque (N•m) 5 repetitions Highest or average value?
Forthomme et al[25]	N=12 Condition: Healthy subjects Age: 24.4±1.7 y Sex: 12 male	Cybex	Time test-retest: 10 d Position: Seated, supine Dominant arm Examiners not described	IR+ER Concentric 60°/s+240°/s	Peak torque (N•m) 3 or 5 repetitions Highest or average value?
Frisiello et al[26]	N=18 Condition: Healthy subjects Age: 18-30 y Gender: 6 male, 12 female	Biodex	Time test-retest: 7 d Position: Standing Bilateral One examiner	IR+ER Eccentric 90°/s+120°/s	Peak torque (N•m) 3 repetitions Highest or average value?
Grabowski et al[27]	N=44 Condition: Healthy subjects Age: 21.2±1.5 y Sex: 24 male, 20 female	Cybex	Time test-retest: 1 min Position: Standing Dominant arm Examiners not described	ER Isometric	Peak torque (N•m) converted to force (N) 3 repetitions
Habets et al[6]	N=49 Condition: Healthy subjects Age: 20.9±3.1 y Sex: 25 male, 24 female	Humac NORM	Time test-retest: 1-2 wk 2 different devices Position: Supine Bilateral 1 examiner, physiotherapy student with limited experience	IR+ER Concentric: IR+ER Eccentric: ER 60°/s+120°/s	Peak torque (N•m) 3 or 5 repetitions Highest value
Hill et al[28]	N=17 Condition: Healthy subjects Age: 28.4±7.0 y Sex: 11 male, 6 female	Cybex	Time test-retest: 7 d Position: Seated, supine Bilateral 1 examiner	IR+ER Concentric+isometric 60°/s+90°/s+120°/s	Peak torque (N•m) 5 repetitions Highest or average value?

(*continued on next page*)

Table 1 (*continued*)

Author	Participants	Type of Dynamometer	Procedure	Movement/Velocity	Unit Measurement
Kramer et al[29]	N=40 Condition: Healthy subjects Age: 58.5±9.0 y Sex: 20 male, 20 female	Kinetic Communicator	Time test-retest: 4-6 d Position: Seated Dominant arm Examiners not described	IR+ER+ER/IR ratio Isometric+concentric+eccentric 60°/s+120°/s	Peak torque (N•m) 2 repetitions Average value
Leggin et al[30]	N=17 Condition: Healthy subjects Age: 30±6 y Sex: 7 male, 10 female	Biodex	Time test-retest: Interrater: 30 min Intrarater: Within 1 wk Position: Seated Abd, standing IR, ER Right arm 2 examiners	IR+ER+Abd Isometric	3 repetitions Within session intrarater reliability: comparing 3 repetitions in 1 session Interrater reliability: Comparing 2 repetition
Lindstrøm et al[31]	N=27 Condition: Healthy subjects Age: Female 35±5 y, male 38±7 y Sex: 13 male, 14 female	KinCom	Time test-retest: 1 wk Position: Seated Right arm 1 examiner	Flex Isometric+concentric 30°/s+90°/s	Isometric: Peak force (N) Concentric: Highest mean force value of 3 repetitions (N) 3 repetitions Highest value
Magnusson et al[32]	N=9 Condition: Healthy subjects Age: 22-34 y Sex: 4 male, 5 female	Cybex	Time test-retest: 1-2 wk Position: Seated Dominant or nondominant arm 1 examiner	Abd+Add Concentric 60°/s	Peak torque (N•m) 6 repetitions Average value
Malerba et al[33]	N=24 Condition: Unilateral shoulder pathology Age: 17-58 y Sex: 14 male, 10 female	Biodex	Time test-retest: 1 wk Position: Seated Bilateral 1 examiner	IR+ER Concentric: 60°/s+120°/s Eccentric: 60°/s	Peak torque (N•m) Work (J) Average power (W) 3 repetitions Highest or average value?
Mandalidis et al[34]	N=31 Condition: Healthy subjects Age: 23 y; 18-34 y Sex: 31 male	KinCom	Time test-retest: 6-7 d Position: Seated Bilateral 1 examiner, experienced physiotherapist Gravity corrected	IR+ER Concentric+eccentric 60°/s+120°/s	Peak torque (N•m) Average moment (N•m) 3 repetitions Highest and average value
Mayer et al[35]	N=29 Condition: Healthy subjects Age: 20-45 y Sex: 14 male, 15 female	Lido	Time test-retest: 2 wk Position: Supine Bilateral Examiners not described	IR+ER+Flex+Ex+Abd+Add Isometric Concentric: 60°/s+180°/s+240°/s+300°/s Eccentric: 60°/s+120°/s+180°/s+240°/s	Peak torque (N•m) Angle at peak torque 5 repetitions Mean of the 3 highest values

(*continued on next page*)

Table 1 (*continued*)

Author	Participants	Type of Dynamometer	Procedure	Movement/Velocity	Unit Measurement
Meeteren et al[36]	N=20 Condition: Healthy subjects Age, asymmetrical use of arm: 27±9.6 y Age, symmetrical use of arm: 32±12.7 y Sex: 10 male, 10 female	Biodex	Time test-retest: 2 wk Position: Seated Bilateral Gravity corrected 1 or 2 examiners	IR+ER Abd+Add: 60°/s+120°/s ER+IR: 60°/s+180°/s	Peak torque (N•m) 60°/s: 5 repetitions 120°/s+180°/s: 10 repetitions Highest value
Papotto et al[37]	N=10 Condition: Healthy subjects Age: 30±12 y Sex: ? male, ? female	Cybex	Time test-retest: 1 wk Position: Seated Dominant arm Examiners not described	ER Isometric in 7 angles Eccentric: 60°/s	Isometric: 2 repetitions, average of the 2 Eccentric: 6 repetitions, average of the middle 4
Plotnikoff et al[38]	N=14 Condition: Healthy subjects Age: 29.9 y Sex: 6 male, 8 female	KinCom	Time test-retest: 2-21 d Position: Seated Bilateral 1 examiner	IR+ER Concentric+eccentric 30°/s	Torque (N•m) 4 repetitions Average over repetition 2-4
Smith et al[39]	N=10 Condition: Healthy subjects Age: 27.7±6.4 y Sex: 5 male, 5 female	KinCom	Time test-retest: 24-72 h Position: Seated Dominant right arm 2 examiners	IR+ER Isometric (5s)+concentric 90°/s	Peak torque (N•m) 2-3 isometric repetitions 5 concentric repetitions Highest or average value?
Sullivan et al[40]	N=14 Condition: Healthy subjects Age: 23 y Sex: 14 male	Cybex	Time test-retest: 1 wk Position: Supine Dominant arm 1 examiner	ER Isometric	Peak torque (N•m) 2-3 isometric repetitions 5 concentric repetitions Average value

Abbreviations: Abd, abduction; Add, adduction; Flex, flexion; Ex, extension.

Table 2 Rating of results and risk of bias assessment for reliability and measurement error of isokinetic dynamometry

	Reliability		Measurement Error	
Author	Rating of Results	Risk of Bias Assessment	Rating of Results	Risk of Bias Assessment
Anderson et al[21]	+	D	−	A
Dauty et al[22]	+	A	?	A
Cavuoto et al[23]	+	A	+	A
Edouard et al[24]	+	A	−	A
Forthomme et al[25]	?	I	−	A
Frisiello et al[26]	+	D		
Grabowski et al[27]	+	V	−	V
Habets et al[6]	+	A	−	A
Hill et al[28]	?	D		
Kramer et al[29]	?	D	−	D
Leggin et al[30]	+	D		
Lindstrøm et al[31]	?	D	−	V
Magnusson et al[32]	+	D		
Malerba et al[33]	?	D		
Mandalidis et al[34]	+	D	−	A
Mayer et al[35]	?	I		
Meeteren et al[36]	+	I	−	I
Papotto et al[37]	+	D	−	A
Plotnikoff et al[38]	+	D	−	D
Smith et al[39]	+	D		
Sullivan et al[40]	?	D		

Abbreviations: A, adequate; D, doubtful; I, inadequate; −, insufficient; +, sufficient; ?, indeterminate; V, very good.

The most common reasons for downgrading the quality of evidence were risk of bias and imprecision. We downgraded for risk of bias when there were few studies and/or the methodological quality was "doubtful" or "inadequate." Imprecision was downgraded 1 level when the total sample size was <100 and 2 levels when it was <50. Inconsistency was rarely downgraded and indirectness was not downgraded in any of the strata investigated. Very low quality of evidence emerged only if the given stratum was examined in just 1 study with a sample size <50 (downgraded for both risk of bias and imprecision).

The reliability results for ID were overall acceptable. However, for clinicians, measurement error is often more relevant than reliability because they are typically interested in repeating muscle strength assessment in the same individual over time.[8] The measurement error needs to be smaller than the MIC to be able to detect small but clinically relevant changes.[11] The size of MIC for muscle strength is not defined in any official guidelines. Lombardi et al[41] and Bae et al[42] found that changes in muscle strength ranged between 7% and 23% after a strength exercise intervention in a population of patients with shoulder disorders. Other studies indicated a change in muscle strength of 10%-15% as clinically relevant.[4,19] Based on these results, we considered an MIC of 15% as being adequate, and we rated the measurement error results against this criterion. In this review, very few of the reported measurement error results were lower than 15%, and all of the strata investigated were rated as "insufficient." The sensitivity analysis showed that even if MIC was increased to 20% all strata were still rated as insufficient. Therefore, we must conclude that ID cannot measure changes in muscle strength <20%, and evaluation of treatment effect should be interpreted with caution. If clinicians are interested in repeating muscle strength assessment in the same individual over time, the increase or decrease in muscle strength must exceed 20% to be sure that the change in muscle strength is larger than the measurement error of the instrument. Whether this change is clinically meaningful depends on the expected magnitude of change in the population of interest. The expected change is affected by the baseline characteristic of the subjects and the intervention evaluated. ID is considered the criterion standard in muscle strength assessment[1,5]; other instruments with better measurement properties do not exist. However, knowledge of performance of the instrument will help the clinicians to interpret results from research or clinical practice.

Our results are similar to those reported in other reviews even though they had a slightly different focus.[2,3] Edouard et al[3] found the seated position to be the most reliable compared with the standing and supine position. Rabelo et al[2] found ICC rating from 0.87-0.92 and %SEM from 15%-24% when evaluating shoulder flexion and extension.

It is suggested that concentric strength testing should be more reliable than eccentric testing.[21,33,38] Some explanations could be that concentric contractions produce lower forces than eccentric contractions, most of the daily movements of the shoulder are concentric, and people are generally unfamiliar with eccentric work.[34,35] More studies in this review examined the concentric test mode than isometric and eccentric test mode. Eight studies investigated both concentric and eccentric contractions; results were not consistent. Therefore, there is no strong evidence to support that the measurement properties of concentric contractions are better than of eccentric contractions.

Three studies investigated the strength imbalance ER_{con}/IR_{con} ratio and found ICC=0.51-0.80 and %SEM=12%-27%.[22,24,29] Only Edouard et al[24] investigated other ratios than the agonist-antagonist ratios in the concentric test mode, for example, bilateral concentric and eccentric ratios. All of these ratios showed low

Table 3 Summary of findings for the reliability of isokinetic dynamometry reported separately for IR, ER, and ER/IR ratio; for isometric, concentric, and eccentric test mode, for seated, supine, and standing position; and for velocities 30°/s, 60°/s, 90°/s, 120°/s, and 240°/s

Variables	No. of Studies, No. of Participants	Summary of Results ICC (Range)	Summary of Results Proportion of ICC≥0.70 (%)	Overall Rating of Results	Quality of Evidence
IR isometric					
Seated	2 studies, n=27	0.82-0.99	100	Sufficient	Low
Supine	1 study, n=17	0.87-0.89	100	Sufficient	Very low
Standing	1 study, n=17	0.96-0.97	100	Sufficient	Very low
IR concentric					
Seated, 30°/s-60°/s	8 studies, n=176	0.39-0.97	94	Sufficient	Moderate
Seated, 90°/s	2 studies, n=27	0.32-0.98	66	Indeterminate	Low
Seated, 120°/s	5 studies, n=132	0.11-0.98	90	Sufficient	Moderate
Supine, 60°/s	2 studies, n=66	0.86-0.94	100	Sufficient	Moderate
Supine, 90°/s	1 study, n=17	0.87-0.93	100	Sufficient	Very low
Supine, 120°/s	2 studies, n=66	0.88-0.94	100	Sufficient	Moderate
IR eccentric					
Seated, 30°/s-60°/s	4 studies, n=115	0.70-0.96	100	Sufficient	High
Seated, 120°/s	1 study, n=31	0.77-0.86	100	Sufficient	Very low
Standing, 90°/s	1 study, n=18	0.75-0.78	100	Sufficient	Low
Standing, 120°/s	1 study, n=18	0.83	100	Sufficient	Low
ER isometric					
Seated	3 studies, n=37	0.85-0.99	100	Sufficient	Low
Supine	1 study, n=17	0.73-0.88	100	Sufficient	Very low
Standing	2 studies, n=61	0.96-0.97	100	Sufficient	Moderate
ER concentric					
Seated, 30°/s-60°/s	8 studies, n=176	0.70-0.95	100	Sufficient	High
Seated, 90°/s	2 studies, n=27	0.74-0.99	100	Sufficient	Low
Seated, 120°/s	5 studies, n=132	0.62-0.92	80	Sufficient	Moderate
Supine, 60°/s	2 studies, n=66	0.84-0.93	100	Sufficient	Moderate
Supine, 90°/s	1 study, n=17	0.75-0.82	100	Sufficient	Very low
Supine, 120°/s	2 studies, n=66	0.74-0.94	100	Sufficient	Low
ER eccentric					
Seated, 30°/s-60°/s	6 studies, n=139	0.44-0.98	83	Sufficient	Moderate
Seated, 120°/s	2 studies, n=45	0.86-0.96	100	Sufficient	Low
Supine, 60°/s	1 study, n=49	0.78-0.91	100	Sufficient	Very low
Supine, 120°/s	1 study, n=49	0.72-0.80	100	Sufficient	Very low
Standing, 90°/s	1 study, n=18	0.78-0.86	100	Sufficient	Low
Standing, 120°/s	1 study, n=18	0.83	100	Sufficient	Low
ER_{con}/IR_{con}					
Seated, 60°/s	2 studies, n=60	0.50-0.79	75	Sufficient	Moderate
Seated, 120°/s	2 studies, n=60	0.53-0.81	50	Indeterminate	Moderate

reliability and large measurement error, and this study revealed that individual ratio changes must be higher than 70% to be detected as a real change.[24] The lower reliability and larger measurement error of results based on ratios compared with basic strength assessment should be taken into account when interpreting results of ratios.

The finding that the affected shoulder showed higher reliability than the nonaffected shoulder was based on 2 relatively small studies of doubtful methodological quality. Therefore, this finding should be given limited effect.

Study limitations

To our knowledge, this is the first systematic review to summarize available research focusing on measurement properties of ID for assessing shoulder muscle strength. However, some limitations must be noted. Relevant studies may have been missed, although we used an extensive literature search including several databases and additional strategies such as reference checking, and we did not use any language restrictions. Not all non-English language journals are indexed in the databases that we used, so we might have missed relevant studies not identifiable in these databases. Additionally, a search for gray literature was not performed because we considered that the majority of relevant studies would be available through databases. Furthermore, we used the COSMIN methodology to evaluate the measurement properties examined in the included studies. The COSMIN methodology seemed relevant and worked well throughout the process. We have tried to describe all steps of the review process in detail to make the process as transparent as possible. However, using other

Table 4 Summary of findings for measurement error of isokinetic dynamometry reported separately for IR, ER, and ER/IR ratio; for isometric, concentric, and eccentric test mode; for seated and supine position; and for velocities 30°/s, 60°/s, 120°/s, and 240°/s

Variables	No. of Studies, No. of Participants	Summary of Results (%SEM, Range)	Summary of Results Proportion of %SEM≤5.4% (%)	Overall Rating of Results	Quality of Evidence
IR isometric					
Seated	1 studies, n=40	14.0	0	Insufficient	Very low
IR concentric					
Seated, 30°/s-60°/s	7 studies, n=173	5.6-27.8	0	Insufficient	High
Seated, 120°/s	3 studies, n=117	8.4-26.6	0	Insufficient	Moderate
Seated, 240°/s	1 study, n=12	9.4	0	Insufficient	Very low
Supine, 60°/s	2 studies, n=61	7.1-10.7	0	Insufficient	Moderate
Supine, 120°/s	1 studies, n=49	9.8-14.5	0	Insufficient	Very low
Supine, 240°/s	1 study, n=12	8.0-12.3	0	Insufficient	Very low
IR eccentric					
Seated, 30°/s-60°/s	4 studies, n=131	9.2-24.9	0	Insufficient	Moderate
Seated, 120°/s	2 studies, n=71	16.7-24.9	0	Insufficient	Low
ER isometric					
Seated	2 studies, n=50	4.4-14.5	13	Insufficient	Low
ER concentric					
Seated, 30°/s-60 °/s	7 studies, n=173	9.8-26	0	Insufficient	High
Seated, 120°/s	3 studies, n=117	8.3-18.1	0	Insufficient	Moderate
Seated, 240°/s	1 study, n=12	17.4	0	Insufficient	Very low
Supine, 60°/s	2 studies, n=61	6.9-9.2	0	Insufficient	Moderate
Supine, 120°/s	1 studies, n=49	7.3-12.1	0	Insufficient	Very low
Supine, 240°/s	1 study, n=12	7.7-8.1	0	Insufficient	Very low
ER eccentric					
Seated, 30°/s-60 °/s	5 studies, n=141	5.2-21.0	11	Insufficient	Moderate
Seated, 120°/s	2 studies, n=71	12.5-17.9	0	Insufficient	Low
Supine, 60°/s	1 studies, n=49	8.6-11.8	0	Insufficient	Very low
Supine, 120°/s	1 studies, n=49	12.2-13.0	0	Insufficient	Very low
ER_{con}/IR_{con}					
Seated, 60°/s	2 studies, n=86	12.1-25	0	Insufficient	Low
Seated, 120°/s	2 studies, n=86	12.2-27	0	Insufficient	Low

quality assessment tools or other inclusion criteria might have generated different conclusions.

Conclusions

We conclude that the reliability of ID was overall sufficient for all positions, velocities, and modes of strength. The measurement error was insufficient for all movements examined. The change in muscle strength must be larger than 20% to be sure that it exceeds the measurement error of the instrument. Because most studies used the seated position, the velocities 30°/s-60°/s or 120°/s, and the concentric test mode, the quality of evidence was highest for these conditions.

Keywords

Rehabilitation; Shoulder

Corresponding author

Lotte Sørensen, MSc, Department of Physiotherapy and Occupational Therapy, Aarhus University Hospital, Palle Juul-Jensens Blvd 99, DK- 8200 Aarhus 8200, Denmark. *E-mail address:* lotsoere@rm.dk.

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;95:2444-69.
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;23:26-35.
3. Edouard P, Samozino P, Julia M, et al. Reliability of isokinetic assessment of shoulder-rotator strength: a systematic review of the effect of position. J Sport Rehabil 2011;20:367-83.
4. Chamorro C, Armijo-Olivo S, Fuente CDl, Fuentes J, Chirosa LJ. 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 (Wars) 2017;12:359-75.
5. 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;3:472-9.
6. Habets B, Staal JB, Tijssen M, van Cingel R. Intrarater reliability of the Humac NORM isokinetic dynamometer for strength measurements of the knee and shoulder muscles. BMC Res Notes 2018;11:15.

7. Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000;30:1-15.
8. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006;59:1033-9.
9. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res 2005;19:231-40.
10. Kottner J, Gajewski BJ, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS). Int J Nurs Stud 2011;48:659-60.
11. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27:1147-57.
12. 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.
13. Higgins JPT, Green S. Cochrane handbook for systematic reviews of interventions version 5.1.0. Available at: www.handbook.cochrane.org. Accessed March 3, 2019.
14. Schünemann H; GRADE handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Available at: https://gdt.gradepro.org/app/handbook/handbook.html. Accessed March 3, 2019.
15. Sørensen L, Oestergaard LG, van Tulder M, Petersen AK. Measurement properties of objective methods to assess shoulder muscle strength (a systematic review protocol). Phys Ther Rev 2017;22:238-42.
16. 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;18:1115-23.
17. 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;27:1171-9.
18. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34-42.
19. 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.
20. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6.
21. Anderson VB, Bialocerkowski AE, Bennell KL. Test-retest reliability of glenohumeral internal and external rotation strength in chronic rotator cuff pathology. Phys Ther Sport 2006;7:115-21.
22. Dauty M, Delbrouck C, Huguet D, Rousseau B, Potiron-Josse M, Dubois C. Reproducibility of concentric and eccentric isokinetic strength of the shoulder rotators in normal subjects 40 to 55 years old. Isokinet Exerc Sci 2003;11:95-100.
23. Cavuoto LA, Pajoutan M, Mehta RK. Reliability analyses and values of isometric shoulder flexion and trunk extension strengths stratified by body mass index. PLoS One 2019;14:e0219090.
24. Edouard P, Codine P, Samozino P, Bernard PL, Herisson C, Gremeaux V. Reliability of shoulder rotators isokinetic strength imbalance measured using the Biodex dynamometer. J Sci Med Sport 2013;16:162-5.
25. Forthomme B, Dvir Z, Crielaard JM, Croisier JL. Isokinetic assessment of the shoulder rotators: a study of optimal test position. Clin Physiol Funct Imaging 2011;31:227-32.
26. Frisiello S, Gazaille A, O'Halloran J, Palmer ML, Waugh D. Test-retest reliability of eccentric peak torque values for shoulder medial and lateral rotation using the Biodex isokinetic dynamometer. J Orthop Sports Phys Ther 1994;19:341-4.
27. Grabowski P, Narveson M, Siegle S. Reliability, responsiveness, and criterion validity of the kiio Sensor, a new tool for assessment of muscle function. Mil Med 2017;182:26-31.
28. Hill AM, Pramanik S, McGregor AH. Isokinetic dynamometry in assessment of external and internal axial rotation strength of the shoulder: comparison of two positions. Isokinet Exerc Sci 2005;13:187-95.
29. Kramer JF, Ng LR. Static and dynamic strength of the shoulder rotators in healthy, 45- to 75-year-old men and women. J Orthop Sports Phys Ther 1996;24:11-8.
30. 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;5:18-24.
31. Lindström B, Waling K, Sundelin G, Ahlgren C. Test-retest reliability of biomechanical output and subjective ratings of exertion in isometric and isokinetic shoulder forward flexion in healthy subjects. Adv Physiother 2003;5:169-78.
32. Magnusson SP, Gleim GW, Nicholas JA. Subject variability of shoulder abduction strength testing. Am J Sports Med 1990;18:349-53.
33. Malerba JL, Adam ML, Harris BA, Krebs DE. Reliability of dynamic and isometric testing of shoulder external and internal rotators. J Orthop Sports Phys Ther 1993;18:543-52.
34. Mandalidis DG, Donne B, O'Regan M, O'Brien M. Reliability of isokinetic internal and external rotation of the shoulder in the scapular plane. Isokinet Exerc Sci 2001;9:65-72.
35. Mayer F, Horstmann T, Kranenberg U, Rocker K, Dickhuth HH. Reproducibility of isokinetic peak torque and angle at peak torque in the shoulder joint. Int J Sports Med 1994;15(Suppl 1):S26-31.
36. Meeteren J, Roebroeck ME, Stam HJ. Test-retest reliability in isokinetic muscle strength measurements of the shoulder. J Rehabil Med 2002;34:91-5.
37. Papotto BM, Rice T, Malone T, Butterfield T, Uhl TL. Reliability of isometric and eccentric isokinetic shoulder external rotation. J Sport Rehabil 2016;25. 2015-0046.
38. Plotnikoff NA, MacIntyre DL. Test-retest reliability of glenohumeral internal and external rotator strength. Clin J Sport Med 2002;12:367-72.
39. Smith J, Padgett DJ, Kotajarvi BR, Eischen JJ. Isokinetic and isometric shoulder rotation strength in the protracted position: a reliability study. Isokinet Exerc Sci 2001;9:119-27.
40. 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;10:213-7.
41. Lombardi JI, Magri AG, Fleury AM, Da Silva AC, Natour J. Progressive resistance training in patients with shoulder impingement syndrome: a randomized controlled trial. Arthritis Rheum 2008;59:615-22.
42. Bae YH, Lee GC, Shin WS, Kim TH, Lee SM. Effect of motor control and strengthening exercises on pain, function, strength and the range of motion of patients with shoulder impingement syndrome. J Phys Ther Sci 2011;23:687-92.