ACRM
AMERICAN CONGRESS OF REHABILITATION MEDICINE

Archives of Physical Medicine and Rehabilitation

journal homepage: www.archives-pmr.org

Archives of Physical Medicine and Rehabilitation 2014;95:2444-69

REVIEW ARTICLE

Intraexaminer Reliability of Hand-Held Dynamometry in the Upper Extremity: A Systematic Review

Patrick P.M. Schrama, MSc, PT, MPT,[a,b] Martijn S. Stenneberg, MSc, PT, MPT,[b] Cees Lucas, PhD, PT,[c] Emiel van Trijffel, MSc, PT, MPT[b,c]

From the [a]private practice Leidsevaart, Haarlem; [b]Stichting Opleidingen Musculoskeletale Therapie, Institute for Master Education in Musculoskeletal Therapy, Amersfoort; and [c]Department of Clinical Epidemiology, Biostatistics, and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

Abstract

Objective: To summarize and appraise the literature on the intraexaminer reliability of hand-held dynamometry (HHD) in the upper extremity.

Data Sources: MEDLINE, CINAHL, and EMBASE were searched for relevant studies published up to December 2011. In addition, experts were contacted, and journals and reference lists were hand searched.

Study Selection: To be included in the review, articles needed to (1) use a repeated-measures, within-examiner(s) design; (2) include symptomatic or asymptomatic individuals, or both; (3) use HHD to measure muscle strength in any of the joints of the shoulder, elbow, or wrist with the "make" or the "break" technique; (4) report measurements in kilogram, pound, or torque; (5) use a device that is placed between the examiner's hand and the subject's body; and (6) present estimates of intraexaminer reliability.

Data Extraction: Quality assessment and data extraction were performed by 2 reviewers independently.

Data Synthesis: Fifty-four studies were included, of which 26 (48%) demonstrated acceptable intraexaminer reliability. Seven high-quality studies showed acceptable reliability for flexion and extension of the elbow in healthy subjects. Conflicting results were found for shoulder external rotation and abduction. Reliability for all other movements was unacceptable. Higher estimates were reached for within-sessions reliability and if means of trials were used.

Conclusions: Intraexaminer reliability of HHD in upper extremity muscle strength was acceptable only for elbow measurements in healthy subjects. We provide specific recommendations for future research. Physical therapists should not rely on HHD measurements for evaluation of treatment effects in patients with upper extremity disorders.

Archives of Physical Medicine and Rehabilitation 2014;95:2444-69

Muscle strength testing is one of the most common impairment measures used by physical therapists.[1] It provides numeric data about strength grade that have relevance for determining patient function before, during, or after intervention.[2] Various core sets and guidelines recommend to measure muscle strength.[3-8]

Manual muscle testing is one of the most commonly used methods for assessing muscle strength.[1] However, several studies[9-11] concluded that this method fails to differentiate among patients with various degrees of muscle weakness compared with more objective methods such as isokinetic dynamometry and hand-held dynamometry (HHD). Although isokinetic dynamometry is considered the criterion standard in measuring muscle strength, upper extremity muscles never generate isokinetic muscle actions during real-life functioning. Moreover, in daily clinical practice, hand-held dynamometers are portable, small, easy to use, minimally time-consuming, and relatively inexpensive compared with isokinetic dynamometers.[12-15] A systematic review[16] found 19 studies correlating the isokinetic dynamometer with the hand-held dynamometer. It was concluded that HHD for upper extremity muscles was valid for measuring muscle strength.[16]

In HHD, the instrument is grasped in the hand of the examiner.[17] Two techniques of testing have been described in the literature. The "break test" requires that the examiner pushes against a subject's extremity until the subject's maximal muscular effort is overcome, thereby producing an eccentric contraction. In the "make test," the examiner holds the dynamometer stationary while the subject exerts a maximal force against it and produces an isometric

Disclosures: none.

0003-9993/14/$36 - see front matter © 2014 by the American Congress of Rehabilitation Medicine
http://dx.doi.org/10.1016/j.apmr.2014.05.019

contraction.[17] Both techniques can be used but are not interchangeable because the "break test" produces higher forces.[18,19]

Objectivity is crucial when assessing muscle strength in a clinical setting and depends on the quality of the measurements.[20] In addition to validity, another requirement of measurements is their reliability. It concerns how well patients can be distinguished from each other, despite measurement error.[20] In physical therapy practice, patients prefer longitudinal continuity of care, which means they receive treatment by 1 provider.[21] Therefore, muscle strength measurements should be reliable when 1 examiner repeatedly evaluates the same patient. This source of error can be estimated by the intraexaminer reliability.[22] Kolber and Cleland[23] summarized the intraexaminer reliability of HHD for both healthy participants and patients with known impairments or disorders. In this narrative review, they concluded that HHD was reliable within examiners for measuring upper extremity muscle strength. However, no study to systematically identify, select, critically appraise, and analyze studies on HHD has been conducted.[24] We conducted a systematic review on the intraexaminer reliability for HHD in the upper extremity.

Methods

Search strategy

MEDLINE (PubMed), CINAHL, and EMBASE were searched for studies published between 1966, 1980, and 1982, respectively, and December 2011. The search strategy was developed with the help of an experienced clinical librarian. Medical Subject Heading (MeSH) terms included "Muscle Strength Dynamometer," "Muscle, Skeletal/physiology," "Extremities," "Isometric Contraction," "Muscle Weakness," "Muscle Strength," "Joints," "Observer Variation" and "Reproducibility of Results" supplemented with free text words and synonyms. The MEDLINE search was translated to EMBASE and CINAHL (appendices 1–3). In addition, reference lists of all retrieved articles were hand searched for relevant studies by 1 reviewer (P.S.). Furthermore, 5 journals frequently publishing articles involving HHD were hand searched including *Journal of Orthopaedic & Sports Physical Therapy, Archives of Physical Medicine and Rehabilitation, Physical Therapy, Isokinetic and Exercise Science,* and *Physiotherapy Canada* (January 1983 to December 2011). Finally, 21 experts were contacted asking if they could provide additional studies.

Study selection

Titles and abstracts were screened by 2 reviewers (P.S., M.S.) independently. When relevant and published as an original full text, articles were obtained and included if they met the inclusion criteria. Studies included met the following criteria:

1. *Design:* Repeated measures within examiner(s).
2. *Participants:* Symptomatic and asymptomatic individuals.
3. *Measurement procedure:* Measurements using a hand-held dynamometer to measure muscle strength in any of the joints of the shoulder, elbow, or wrist with the "make" or the "break" technique. The device must be placed between the examiner's hand and the subject's body. The strength is reported in kilograms, pounds, newtons, or torque.
4. *Outcome:* Estimates of intraexaminer reliability.

Devices measuring muscle tone or developed to measure just 1 joint were excluded. Articles assessing patients with central neurologic disorders or intellectual developmental disorders were excluded because reliability estimates could be influenced substantially by muscle tone, medication, or motivational problems. Studies using isokinetic devices or fixed hand-held dynamometers were also excluded. In fixed HHD, a system is needed to fix the hand-held dynamometer, which might affect its portability and cost-effectiveness.[25] Abstracts and documents that were speculative, anecdotal, or editorial in nature were not considered. When the full-text article was not available after contacting the authors, the study was excluded. No restrictions were made on language or date of publication. If disagreement persisted, a third reviewer (C.L.) had a final judgment.

List of abbreviations:

CI	**confidence interval**
COPD	**chronic obstructive pulmonary disease**
HHD	**hand-held dynamometry**
ICC	**intraclass correlation coefficient**
QAREL	**Quality Appraisal of Diagnostic Reliability Studies**
QUADAS-2	**Quality Assessment of Diagnostic Accuracy Studies**

Quality assessment

Criteria for assessing methodological quality of included studies were derived from the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2),[26] Standards for the Reporting of Diagnostic Accuracy studies,[27,28] Quality Appraisal of Diagnostic Reliability Studies (QAREL),[29] and existing quality checklists used in systematic reviews of diagnostic reliability.[30-34] QUADAS-2 and QAREL were partly inappropriate in their original format. QUADAS-2 aims to evaluate studies of diagnostic accuracy in which a test is used with the purpose of detecting or predicting a target condition and to formulate a diagnosis.[34] In assessing muscle strength, there is no disease status that needs to be diagnosed. Several items in the QAREL tool are not applicable for intraexaminer reliability of strength measurements.[29]

We compiled a list of 11 quality criteria (table 1). Each criterion is answered with "yes," "no," or "unclear" when insufficient information was provided. Criteria 1 through 6 were used to support the risk of bias judgment—that is, the internal validity of a study where "yes" indicated a low risk of bias. A study was considered to have a low risk of bias if $\geq$5 criteria were rated as "yes," a moderate risk if 3 or 4 criteria were rated as "yes," and a high risk if $\leq$2 criteria were rated as "yes." It is suggested that scoring "no" to criteria 1 through 3 and 6 may reflect overestimated reliability, while scoring "no" to criteria questions 4 and 5 may reflect underestimated reliability.[26,35] Criteria 7 through 10 concern the applicability of results or external validity of a study. Applicability was considered sufficient if all criteria were scored as "yes." A detailed list of instructions was developed for each item and is available from the first author (P.S.). The criteria were first tested in a training session. Consensus was achieved after several rounds of discussions among the authors. Finally, 2 reviewers (P.S., M.S.) independently assessed the quality of all included studies but were not blinded to information on authors and journal. Disagreements were resolved by discussion among the reviewers. If disagreement persisted, a third reviewer (C.L.) had a final judgment.

Data collection

Data collection was performed by 2 reviewers (P.S., M.S.) independently using an electronic data extraction form. Data on subjects and examiners, measurement procedure, muscle function of